

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 14:39:32 ; Search time 539.84 Seconds

(without alignments)
135.498 Million cell updates/sec

Title: US-09-325-095-12
Perfect score: 23
Sequence: 1 AARATGATGATGATHTTTCATAC 23

Scoring table:

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl: *
1: gb_dal: *
2: gb_ba2: *
3: gb_cm: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pl3: *
10: gb_pl4: *
11: gb_pl5: *
12: gb_pl6: *
13: gb_pl7: *
14: gb_pl8: *
15: gb_pl9: *
16: gb_pl10: *
17: gb_pl11: *
18: gb_pl12: *
19: gb_pl13: *
20: gb_pl14: *
21: gb_pl15: *
22: gb_pl16: *
23: gb_pl17: *
24: gb_pl18: *
25: gb_pl19: *
26: gb_pl20: *
27: gb_pl21: *
28: gb_pl22: *
29: gb_pl23: *
30: gb_pl24: *
31: gb_pl25: *
32: gb_pl26: *
33: gb_pl27: *
34: gb_pl28: *
35: gb_pl29: *
36: gb_pl30: *
37: gb_pl31: *
38: gb_pl32: *
39: gb_pl33: *
40: gb_pl34: *
41: gb_pl35: *
42: gb_pl36: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	20.6	89.6	3207 3	BOVPHOS3KN M93252 Bovine phop

2	20.6	89.6	3452	4	AF001076
3	20.6	89.6	3498	5	A37232
4	20.6	89.6	23	5	A37240
5	20.6	89.6	3424	10	HSPH13K
6	20.6	89.6	3207	10	HSU79143
7	20.6	89.6	3207	12	MM03279
8	20.6	89.6	3389	17	AF001075
9	18.6	80.9	5292	1	SEU77778
10	18	78.3	310617	34	HS69B13
11	18	78.3	5392	36	DMJ12634
12	18	78.3	5381	37	DMJ12634
13	17.8	77.4	199046	35	AC007744
14	17.4	75.7	108878	11	AC003092
15	17.4	75.7	259474	11	HUAC004605
16	17	73.9	4373	1	SGU61158
17	17	73.9	1777	7	LEPLEC14B
18	17	73.9	209618	10	HS54B20
19	17	73.9	169765	11	AC006075
20	17	73.9	2889	12	RATBEM1
21	17	73.9	4871	12	RNU28938
22	17	73.9	4524	36	CELC46E2
23	17	73.9	30250	36	CELC46E2
24	17	73.9	124170	42	AC006000
25	17	73.9	193772	42	AC007073
26	16.8	73.0	82347	7	AB028614
27	16.8	73.0	140679	9	HS246H3
28	16.8	73.0	59765	11	AC005179
29	16.8	73.0	1000	14	HS246H3
30	16.8	73.0	186086	35	AC007165
31	16.8	73.0	184269	35	AC007271
32	16.8	73.0	81053	37	AC005762
33	16.6	72.2	251124	11	HUAE000660
34	16.4	71.3	1359	1	CRY16137
35	16.4	71.3	40245	1	MLCB2533
36	16.4	71.3	40245	1	MTCV261
37	16.4	71.3	27322	1	MTCV261
38	16.4	71.3	42157	1	U00017
39	16.4	71.3	6429	2	AF088800
40	16.4	71.3	40892	5	A59672
41	16.4	71.3	94301	7	ATF28A21
42	16.4	71.3	7074	7	SCYUR152M
43	16.4	71.3	2354	7	YSCDRL5A
44	16.4	71.3	86004	9	HS51723
45	16.4	71.3	103933	10	HS684024

ALIGNMENTS

RESULT 1	BOVPHOS3KN	3207 bp	MRNA	MAM
LOCUS	Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete cds.			
DEFINITION	BOVPHOS3KN 3207 bp			
ACCESSION	M93252			
NID	G163519			
VERSION	M93252.1 G1:163519			
KEYWORDS	Phosphatidylinositol 3-kinase.			
SOURCE	Bos taurus CDNA to mRNA.			
ORGANISM	Bos taurus			
REFERENCE	1			
AUTHORS	Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R., Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F., Hsuan, J.J., Courtemanche, S.A., Parker, P.J., and Waterfield, M.D.			
TITLE	Phosphatidylinositol 3-kinase: Structure and expression of the 110 kD catalytic subunit			
JOURNAL	Cell 70, 419-429 (1992)			
MEDLINE	92354059			
FEATURES	Location/Qualifiers			
SOURCE	1..3207			

CDS

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBAR-1"
1..3207
/EC_number="2.7.1.67"
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="AA30698.1"
/db_xref="GI:163520"
/db_xref="GI:163520"
/translation="MPPRPSGELMGHIMLPRIIVCELLPNGMIVTTECLREATLIT
IKHELFKARKYPRYLQIQDSSSYTFVSVOEAREEPEFDETRLCDDLLPPLKVI
EVLGNREKILNREIGFALGMPECEMDKPEVODFRRNILLNCKEADVLIDNSHP
SRAMYYPRNVSSELPKRIITNKDKQIIVIVIVSPNDKQYTKIKINDCYPE
OYAEAIRKTRSMILSEQLKCYLEVGKYLKVCQDEYFLEKYPLOSKYIRSC
IMLGRPMIMLAKESLSQLEKQVETGKYSRISTATPYNGEYSLTSLWINSAL
RIKILCAIVVNIIDIKIYVATGIVHGEPLCDNVTPVPCSNPRNEMLVNDIY
IPDLPRARILCISYVGRKGAKEHCPPLAGNINLFDYTDIVSGKALMLAPVPH
GLEDLNPIGVTSNPNKEPCLLEEDWESNPVAFPDVTIIEHANVTISREGFNT
SVAGLSNRLADNENLESDKQDLRAICTDPISETTEQKDFLMSHRHCVTIPETLP
KLLSVKNSRDEVAQMYCLVDPPIKEQAMLELDNYPDPVWGFVAVKLELYLT
DKLSQYLIQIVQVLYKEQYLDNOLVRELKALNORIGHFEMHLKSEHNKTVSQ
RGLLEESYCRACGMYLKLNROVEAMEKILNLDILKOEKDETOKVOMKFLVQMR
RDPEDMALQGFSPINPAHQNLNLEECRIMSARPLMLNEMNDIMSELQONE
IIFKNGDDLRODMLTQITRIMENIMQNGDLRMLPGCLSIGCVGLIYVNSHT
IMQIOCKGKALQALQNSHTLHOMLKDKRKGEIYDAIDILFTRSCAGYVATPIIGIG
DHNNSINWKGQGFHIDFGFHLDKKKFGYKREKRVFVLTPDPLIVISGADECT
KTRERFQEMCYKAYLAIROHANLEFINFSMLGSGMPELOSFDIAYIRKTLADK
TEOEALVEYFOMKMDAHHGWTCKMDWIFHTIKOHALN"
BASE COUNT      1028 a      581 c      680 g      918 t
ORIGIN

```

Query Match
Best Local Similarity 78.3%; Score 20.6; DB 3; Length 3207;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 AARATGAYTGATHTTTCAYAC 23
11:111111:111111:111111:111111

RESULT 2 AF001076 3452 bp mRNA VRT 08-JUL-1997
LOCUS Gallus gallus phosphoinositide 3-kinase catalytic subunit mRNA,
DEFINITION complete cds.
ACCESSION AF001076
NID 92245505
VERSION AF001076.1 GI:2245505
KEYWORDS
SOURCE Chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 3452)
AUTHORS Chang, H.W., Aoki, M., Furman, D., Auger, K.R., Bellacosa, A.,
Tsichlis, P.N., Cantley, L.C., Roberts, T.M. and Vogt, P.K.
TITLE Transformation of chicken cells by the gene encoding the catalytic
subunit of PI 3-kinase
JOURNAL Science 276 (5320), 1848-1850 (1997)
MEDLINE 97334438
RECORD 2 (bases 1 to 3452)
AUTHORS Chang, H.W., Aoki, M. and Vogt, P.K.
TITLE Direct Substitution
JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
BCC-239, La Jolla, CA 92037, USA

FEATURES
source 1..3452
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="Brain"

5'UTR
gene

CDS

```

1..150
151..3354
/gene="c-p3k"
/note="p3k proto-oncogene"
151..3357
/note="c-p3k"
/codon_start=1
/product="phosphoinositide 3-kinase catalytic subunit"
/protein_id="AAB62534.1"
/db_xref="PID:92245506"
/db_xref="GI:2245506"
/translation="MPPRPSGELMGHIMLPRIIVCELLPNGMIVTTECLREATLIT
IKHELFKARKYPRYLQIQDSSSYTFVSVOEAREEPEFDETRLCDDLLPPLKVI
EVLGNREKILNREIGFALGMPECEMDKPEVODFRRNILLNCKEADVLIDNSHP
SRAMYYPRNVSSELPKRIITNKDKQIIVIVIVSPNDKQYTKIKINDCYPE
OYAEAIRKTRSMILSEQLKCYLEVGKYLKVCQDEYFLEKYPLOSKYIRSC
IMLGRPMIMLAKESLSQLEKQVETGKYSRISTATPYNGEYSLTSLWINSAL
RIKILCAIVVNIIDIKIYVATGIVHGEPLCDNVTPVPCSNPRNEMLVNDIY
IPDLPRARILCISYVGRKGAKEHCPPLAGNINLFDYTDIVSGKALMLAPVPH
GLEDLNPIGVTSNPNKEPCLLEEDWESNPVAFPDVTIIEHANVTISREGFNT
SVAGLSNRLADNENLESDKQDLRAICTDPISETTEQKDFLMSHRHCVTIPETLP
KLLSVKNSRDEVAQMYCLVDPPIKEQAMLELDNYPDPVWGFVAVKLELYLT
DKLSQYLIQIVQVLYKEQYLDNOLVRELKALNORIGHFEMHLKSEHNKTVSQ
RGLLEESYCRACGMYLKLNROVEAMEKILNLDILKOEKDETOKVOMKFLVQMR
RDPEDMALQGFSPINPAHQNLNLEECRIMSARPLMLNEMNDIMSELQONE
IIFKNGDDLRODMLTQITRIMENIMQNGDLRMLPGCLSIGCVGLIYVNSHT
IMQIOCKGKALQALQNSHTLHOMLKDKRKGEIYDAIDILFTRSCAGYVATPIIGIG
DHNNSINWKGQGFHIDFGFHLDKKKFGYKREKRVFVLTPDPLIVISGADECT
KTRERFQEMCYKAYLAIROHANLEFINFSMLGSGMPELOSFDIAYIRKTLADK
TEOEALVEYFOMKMDAHHGWTCKMDWIFHTIKOHALN"
BASE COUNT      1073 a      672 c      783 g      924 t
ORIGIN

```

Query Match
Best Local Similarity 78.3%; Score 20.6; DB 4; Length 3452;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 AARATGAYTGATHTTTCAYAC 23
11:111111:111111:111111:111111

RESULT 3 A37232 3498 bp DNA PAT 05-MAR-1997
LOCUS Sequence 1 from Patent WO9403609.
DEFINITION A37232
ACCESSION A37232
NID 92284345
VERSION A37232.1 GI:2294345
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 3498)
AUTHORS Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
TITLE EUCARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 1 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
source 1..3498
Location/Qualifiers

1..3207
/organism="unidentified"
/db_xref="taxon:32644"
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PID:e304808"
/db_xref="PID:92294346"
/db_xref="GI:2294346"


```

/translation="MPRSSGELGHLMPRIIVECLIPNGMIVTLECLREATLIT
IKHELFEARKRYPLHQLQDESSYIFVSTOAEEREFDETRRLCDLRLPPLKVI
EPVGRREKILNREIGFALGMPVCEFDVMDKDEVDQFERNILNVEKAVDLNDSPE
SRAMVYPPNVESSPELPRKHYLNLDKQIIVYVIVYSPNNDKOKYTLKINHOCVPE
QVIAIRKTRKSMILSSEOLKICVLEQKILYVCGDEFLERKYPISQYKIRSC
IMLGPMILMAKESLSQLEPDCFTMPYSRISTATPYNGESTRSLAVINSAL
RIKICATVYVNNIRIDIKIYVRGIVYGGGELCGNVMNQRPSCNPRMNEANTDIY
IPDIPRARILCISCSYGRKAKREHCPPLANGNINREDYTTLYSGKALNLYNPPH
GLEDLNLITGTSNPNKETPELEFEMFESSVAFPDMSVIEEHNANSSVSEAGFSY
SHAGLSNRLARNENRENDKEQLAICTRDPLSEITEQEKDLMSHRHYCVTIPILP
KLLSVKNNSRDEVQMYCLVNDMPRIPEQAMELDCNYPDPYRGHAGVCLIEVLT
DKLSQVLIQVQVLYEQYLDNLVRELLKALNORIGHEFFWHLKSEMKNVTSO
REGLLSEYCRACGMYLKHNRQVAMEKILNLTLLQOEKDEOKYOMKRLVYOMR
RPPMDALQGLSPINPAHOLGNLLEERINSKAKRPLMNPMDISELHFOVNE
IIFKGGDLRQDMTLQITRIMENIWNQGDRLRLPGLCSIGCVGLEEVNRSHT
IMQICKGGLKALQFNSHTLHOMKDKNKGIDLYNAIDLFTRSCAGVAFVILIGIG
DRHNSNINVKDQGLFHDGFHLDHKKKKGKREKREPERVUTODFLVYISGAGDECT
KTRERERQEMCYKAVLAIRQHANLFINLSMGLSGMPELQSPFDIAYIRKTLALDK
TEQEALEFYMOMDANHGGWTKMDWIFHTIKOHALN"
BASE COUNT      1147 a      626 c      722 g      1003 t
ORIGIN

Query Match      89.6%; Score 20.6; DB 5; Length 3498;
Best Local Similarity 78.3%; Pred. No. 2.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 AARATGAYTGATTTTCAYAC 23
||:||||:||||:||||:||||:
Db 3160 AAATGATTCATCTCCACAC 3182

RESULT 4
A37240      23 bp      DNA      PAT      05-MAR-1997
LOCUS
DEFINITION Sequence 9 from Patent WO9403609.
ACCESSION A37240
VERSION 92284352
KEYWORDS A37240.1 GI:2294352
SOURCE
ORGANISM unidentified.
unidentified.
unclassified.
REFERENCE 1 (bases 1 to 23)
AUTHORS Parter,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
TITLE EUCARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
PATENT: WO 9403609-A 9 17-FEB-1994;
JOURNAL IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
source 1..23
Location/Qualifiers
BASE COUNT      7 a      2 c      4 g      5 t      5 others
ORIGIN

Query Match      89.6%; Score 20.6; DB 5; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AARATGAYTGATTTTCAYAC 23
|||||
Db 1 AARATGAYTGATTTTCAYAC 23

RESULT 5
HSPH13K      3424 bp      mRNA      PRI      24-AUG-1995
LOCUS
DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
ACCESSION 229090
VERSION 9472990
KEYWORDS

```

```

VERSION 229090.1 GI:472990
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1068)
AUTHORS Rocchi,M. and Waterfield,M.D.
TITLE Molecular cloning, cDNA sequence, and chromosomal localization of
JOURNAL the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene
MEDLINE 95229146
REFERENCE 2 (bases 1 to 3424)
AUTHORS Volinia,S.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig
Institute for Cancer, Research, 91 Riding House Street, London, W1P
8BT, UK

FEATURES
source Location/Qualifiers
1..3424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="K562"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CAA8333.1"
/db_xref="pid:9472991"
/db_xref="GI:472991"
/db_xref="SWISS-PROT:P42336"
/translation="MPRSSGELGHLMPRIIVECLIPNGMIVTLECLREATLIT
IKHELFEARKRYPLHQLQDESSYIFVSTOAEEREFDETRRLCDLRLPPLKVI
EPVGRREKILNREIGFALGMPVCEFDVMDKDEVDQFERNILNVEKAVDLNDSPE
SRAMVYPPNVESSPELPRKHYLNLDKQIIVYVIVYSPNNDKOKYTLKINHOCVPE
QVIAIRKTRKSMILSSEOLKICVLEQKILYVCGDEFLERKYPISQYKIRSC
IMLGPMILMAKESLSQLEPDCFTMPYSRISTATPYNGESTRSLAVINSAL
RIKICATVYVNNIRIDIKIYVRGIVYGGGELCGNVMNQRPSCNPRMNEANTDIY
IPDIPRARILCISCSYGRKAKREHCPPLANGNINREDYTTLYSGKALNLYNPPH
GLEDLNLITGTSNPNKETPELEFEMFESSVAFPDMSVIEEHNANSSVSEAGFSY
SHAGLSNRLARNENRENDKEQLAICTRDPLSEITEQEKDLMSHRHYCVTIPILP
KLLSVKNNSRDEVQMYCLVNDMPRIPEQAMELDCNYPDPYRGHAGVCLIEVLT
DKLSQVLIQVQVLYEQYLDNLVRELLKALNORIGHEFFWHLKSEMKNVTSO
REGLLSEYCRACGMYLKHNRQVAMEKILNLTLLQOEKDEOKYOMKRLVYOMR
RPPMDALQGLSPINPAHOLGNLLEERINSKAKRPLMNPMDISELHFOVNE
IIFKGGDLRQDMTLQITRIMENIWNQGDRLRLPGLCSIGCVGLEEVNRSHT
IMQICKGGLKALQFNSHTLHOMKDKNKGIDLYNAIDLFTRSCAGVAFVILIGIG
DRHNSNINVKDQGLFHDGFHLDHKKKKGKREKREPERVUTODFLVYISGAGDECT
KTRERERQEMCYKAVLAIRQHANLFINLSMGLSGMPELQSPFDIAYIRKTLALDK
TEQEALEFYMOMDANHGGWTKMDWIFHTIKOHALN"
BASE COUNT      1134 a      618 c      709 g      963 t
ORIGIN

Query Match      89.6%; Score 20.6; DB 10; Length 3424;
Best Local Similarity 78.3%; Pred. No. 2.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 AARATGAYTGATTTTCAYAC 23
|||||
Db 3172 AAATGATTCATCTCCACAC 3194

RESULT 6
HSU79143      3207 bp      mRNA      PRI      02-JAN-1997
LOCUS
DEFINITION Human phosphoinositide 3'-hydroxykinase p110-alpha subunit mRNA.
ACCESSION 079143
VERSION 91763625
KEYWORDS 079143.1 GI:1763625

```

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 3207)
TITLE Stridivanti, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledger, L.M., Cliffe, A. and Helmbrook, D.C.
Catalytic activity of the p110-alpha subunit of Human Phosphoinositide 3'-Hydroxylase is Required for Signal Transduction
JOURNAL Bioorg. Med. Chem. (1996) In press
REFERENCE 2 (bases 1 to 3207)
AUTHORS Stridivanti, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledger, L.M., Cliffe, A. and Helmbrook, D.C.
Direct Submission
TITLE Submitted (21-NOV-1996) Cancer Research, Merck Research Labs, Summeytown Pike, West Point, PA 19486, USA
LOCATION/Qualifiers
FEATURES
source 1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS 1..3207
/function="phospholipid kinase"
/product="phosphoinositide 3'-hydroxylase p110-alpha subunit"
/protein_id="AA39753.1"
/db_xref="pid:g1763626"
/db_xref="gi:1763626"
/translation="MPPRSSGELGNGIHMPRLIVCELLPNGMIVTECLREAVLT
IKHELFREARKYPLHQLDSESYTFVSTQAESEDEFDERRLCDLRFQPLKVI
EPVGNREKILNREIGFVIGMPVCEFDVQDEVDPRNLIINCKEAVDLNDSVH
SRAMVYPPNVSSEPLPKRHYIKLKGQITIVIVIVSPNDKQYTKLINDCYPE
QVIAEIRKTRSMILSSBQLKCVLEQKYLKVGCGDEYFLKPYQYKIRSC
IMGRMNPMLMAKESLYSQLPDCETMPSRISTATPYNGESTSLIVYNAL
RIKILCATYVNNIRIDIKYVRIYHGGELPCDNVTPRCSPNRMENLVNDIY
IPDLPRARCLSTISGVKGRKAKHECPPLAMGNLFDYDTLVSGKALNMPVPH
GLEDNLPIGVTSNPNKPCLEEFDFWSSVYKPPMSVIEHANMSYRAGSY
SHAGLSNRLANDNELENDKEQKALSTRDPLSETTBOERDILMSIRHCVTIPILP
KLLSVKMSRDEVAQMYCLVDMPPKPEQAMELLDCNPDVVRGFAVRCLEKYL
DKLSQYLQVAVLYEQLDLMLVRLKALNORIGHFEFHLKSMHAKVTSO
RFGLEISYCRACGMVTLKHLNROVAMKILNIDILKQEKDETOKMKEFLVEQMR
RPFNDALOGFSLPNPAHQGLRLCEGRINSARPLMLNENPDILFQNE
IFKNGDDLRQDMLTQIIRIMENINOGDLRLMPCGLSTGCVGLTEVYRNSHT
IMQIOCKGLKGLKLOENSHRLHQLDKNGEYDAIDLFTRSCAGYVATFIIIG
DRHNSINWAKDQGLFHDGFLDHRKKKFGKRREVPVLTQDFLVIYSKAGQCT
KTRFEFERQEMCYKAYLAIRQANLFINLSMGLSGMPELOSFDIAYIRKTLADK
TQEQLALEYFKQNDNAHGGWTKMDMIFHTIKQNALN"

BASE COUNT 1043 a 586 c 670 g 908 t
ORIGIN

Query Match 89.6%; Score 20.6; DB 10; Length 3207;
Best Local Similarity 78.3%; Pred. No. 2.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 AARATGATGATGATHTTTCAYAC 23
||:||||:||||:||||:||||:||||:
Db 3160 AAAATGATGATGATCTTCACAC 3182

RESULT 7
LOCUS MM003279 3207 bp mRNA ROD 25-MAY-1994
DEFINITION Mus musculus Balb/c phosphatidylinositol 3-kinase 110 kDa subunit
ACCESSION U03279
NID 9414994
VERSION U03279.1 GI:414994
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 3207)
TITLE Kippel, A., Escobedo, J.A., Hirano, M. and Williams, L.T.
The interaction of small domains between the subunits of phosphatidylinositol 3-kinase determines enzyme activity
JOURNAL Mol. Cell. Biol. 14, 2675-2685 (1994)
MEDLINE 9418738
REFERENCE 2 (bases 1 to 3207)
AUTHORS Kippel-Giese, A.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1993) Anke Kippel-Giese, CVRI and Daiichi Research Center, University of California San Francisco, 505 Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
LOCATION/Qualifiers
FEATURES
source 1..3207
/organism="Mus musculus"
/strain="Balb/c"
/db_xref="taxon:10090"
CDS 1..3207
/standard_name="p110 protein"
/codon_start=1
/product="phosphatidylinositol 3-kinase 110 kDa subunit"
/protein_id="AA18334.1"
/db_xref="pid:g414995"
/db_xref="gi:414995"
/translation="MPPRSSGELGNGIHMPRLIVCELLPNGMIVTECLREAVLT
IKHELFREARKYPLHQLDSESYTFVSTQAESEDEFDERRLCDLRFQPLKVI
EPVGNREKILNREIGFVIGMPVCEFDVQDEVDPRNLIINCKEAVDLNDSVH
SRAMVYPPNVSSEPLPKRHYIKLKGQITIVIVIVSPNDKQYTKLINDCYPE
QVIAEIRKTRSMILSSBQLKCVLEQKYLKVGCGDEYFLKPYQYKIRSC
IMGRMNPMLMAKESLYSQLPDCETMPSRISTATPYNGESTSLIVYNAL
RIKILCATYVNNIRIDIKYVRIYHGGELPCDNVTPRCSPNRMENLVNDIY
IPDLPRARCLSTISGVKGRKAKHECPPLAMGNLFDYDTLVSGKALNMPVPH
GLEDNLPIGVTSNPNKPCLEEFDFWSSVYKPPMSVIEHANMSYRAGSY
SHAGLSNRLANDNELENDKEQKALSTRDPLSETTBOERDILMSIRHCVTIPILP
KLLSVKMSRDEVAQMYCLVDMPPKPEQAMELLDCNPDVVRGFAVRCLEKYL
DKLSQYLQVAVLYEQLDLMLVRLKALNORIGHFEFHLKSMHAKVTSO
RFGLEISYCRACGMVTLKHLNROVAMKILNIDILKQEKDETOKMKEFLVEQMR
RPFNDALOGFSLPNPAHQGLRLCEGRINSARPLMLNENPDILFQNE
IFKNGDDLRQDMLTQIIRIMENINOGDLRLMPCGLSTGCVGLTEVYRNSHT
IMQIOCKGLKGLKLOENSHRLHQLDKNGEYDAIDLFTRSCAGYVATFIIIG
DRHNSINWAKDQGLFHDGFLDHRKKKFGKRREVPVLTQDFLVIYSKAGQCT
KTRFEFERQEMCYKAYLAIRQANLFINLSMGLSGMPELOSFDIAYIRKTLADK
TQEQLALEYFKQNDNAHGGWTKMDMIFHTIKQNALN"

BASE COUNT 958 a 693 c 734 g 822 t
ORIGIN

Query Match 89.6%; Score 20.6; DB 12; Length 3207;
Best Local Similarity 78.3%; Pred. No. 2.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 AARATGATGATGATHTTTCAYAC 23
||:||||:||||:||||:||||:||||:
Db 3160 AAAATGATGATGATCTTCACAC 3182

RESULT 8
LOCUS AF001075 3389 bp mRNA VRL 08-JUL-1997
DEFINITION Avian sarcoma virus 16 gag-v-phosphoinositide 3-kinase catalytic subunit fusion protein (gag-v-p3k) and retroviral env mRNAs, partial cds.
ACCESSION AF001075
NID 92245502
VERSION AF001075.1 GI:2245502
KEYWORDS Avian sarcoma virus 16.
SOURCE Avian sarcoma virus 16.
ORGANISM Avian sarcoma virus 16
REFERENCE 1 (bases 1 to 3389)
AUTHORS Chang, H.M., Aoki, M., Furman, D., Auger, K.R., Bellacosa, A., Tsichlis, P.N., Cantley, L.C., Roberts, T.M. and Vogt, P.K.

TITLE	Transformation of chicken cells by the gene encoding the catalytic
JOURNAL	subunit of PI 3-kinase
MEDLINE	Science 276 (5320), 1848-1850 (1997)
REFERENCE	97334438
AUTHORS	2 (bases 1 to 3389)
JOURNAL	Chang, H. W., Aoki, M. and Vogt, P. K.
TITLE	Direct Submission
JOURNAL	* Submitted (23-APR-1997) Molecular & Experimental Medicine, The
FEATURES	Scrrips Research Institute, 10550 N. Torrey Pines Road, Malibu-drop
	BC-239, La Jolla, CA 92037, USA
	Location/Qualifiers

ORF5	Location/Qualifiers
source	1..3389
	/organism="Avian sarcoma virus 16"
	/db_xref="taxon:60629"
gene	<1..3252
	/gene="gag-v-p3k"
	/note="fusion gene of gag and v-p3k"
	1..78
misc_feature	/gene="gag-v-p3k"
	/note="derived from gag gene"
	<1..3252
CDS	1..3389

```

/codon_start=1
/product="gag-V-phosphoinositide 3-kinase catalytic
subunit fusion protein"
/protein_id="AAB62533.1"
/db_xref="PIR:92245503"
/db_xref="GI:2243503"
/translation="QQILRTAPSTLTTPGELITIKYVLDRQKASHMLPRLIECLIPNGA
MIVTEJCEPRAETLTIKHELEFKARKYVLYLLODESEYIVSTOAEERPEEFDR
RLCDRLRQPEPLKTVLEVPNGRNEKRLINKIGFALIPGCEPDMQKDEVDQRRNITR
VCKEAVDLRDANAHPSRALYYCPVWSESPPELPHRIYTKLDKGIIVYIVTYSNND
KQKTLKLNHCVCPEQVIAELAIRKTRTSRNSLSDKLCVLEYGKRIKLVQKDEYL
KQKTLPSQYKIRCSIMGLRMPNLMAKSELTYQLPDTITPMSRSRISTAPYNNM
GEATKSTMTISALIRILCATYANNVIRIDIKRYVYTGHGSEPLCDVNNRQVDP
CSNPWMNMLISWYDIRIPDLPRARALCTISGVKGRKRGKCHCLANGINIMPEYTLT
LYSGMGLINIAVAPGLEDLNLRIIDGVSNNPKRCEOLELEDFMSNPVAKPDMYVIE
EHANNTISRELEFNTYAGLSNRILARDDELRESDEQKRAICTGQDASEITVQKDEL
WSHRIYCVKTEPELPRKLLISVAKMSRDEAVQMYCLVAKMPIKRQDAEMLDCYPPD
MYRAVRAVLEKRYVLEFDKLSQYLVLQVLYEEDYLQDLVRELFILKALTNLRIGHEF
FWHLSESNHKNTVSQRFGLLSEYCSACAGTMLHLSRQVEMLEINTLDLROEKDF
ETOKQMEIVDQMRDPMDALOGFTSLPNPAHOLIGLRLRQRIIMSGMSPALVMM
ENPDLMSLELRQNNETIRKNGDLAQDMLTQITRIMENTNQNGDLIRMLPYCGLSI
GDCVGLIENVSSSHTIWQIQCKGGLKGLAQNSHTLHQLMDKNNGKEUYDAIDLFRF
QDAGCVATFLIGIDRHNNSINMYKDCDQLEPHIGFHLIDRKRFKGRVRFVLTLL
OCLLVICGAGOECKTRREFPEQWCKRATLAEQHNLEINFLSNMGLSGPELQS
FDYIAIKTIALDKTEQEALEYFMKQNNDAHGGWTKMDMITHIKQAHNLN"
79      RA

```

```

misc_feature      79..84
                  /gene="gag-v-p3k"
                  /note="sequence of unknown origin"
misc_feature      85..3252
                  /gene="gag-v-p3k"
                  /note="derived from c-p3k gene"
3' UTR            3250..3370
CDS               3371..>3389
                  /codon_start=1
                  /product="retroviral env"
                  /protein_id="AA62533.1"
                  /db_xref="PID:g2245504"
                  /db_xref="GI:2245504"
                  /translation="MARNSS"
BASE COUNT      1071 a      655 c      751 g      912 t
ORIGIN

```

Query Match	89.6%	Score 20.6	DB 17	Length 3389
Best Local Similarity	78.3%	Pred. No. 2,3		
Matches 18	Conservative	5	Mismatches 0	Indels 0
Gaps				0
Oy	1	AAATGCAITGATHTTCATAC	23	
Db	3205	AAATGAGATGATCTTCACAC	3227	

RESULT	9
LOCUS	SEU77778
DEFINITION	SEU77778 5292 bp DNA BCT 14-JUN-1997
ACCESSION	Staphylococcus epidermidis plasmid pUE2 putative ABC transporter subunits (epig), (epie), and (epit), putative membrane protein (epih), EpiT' (epiT') and EpiT'' (epiT'') genes, complete cds.
NID	U77778 U29130
VERSION	G2196512
KEYWORDS	U77778.1 GI:2196512
SOURCE	.
ORGANISM	Staphylococcus epidermidis. Staphylococcus epidermidis.

REFERENCE AUTHORS TITLE	JOURNAL MEDLINE AUTHORS TITLE
1 (bases 1 to 2600) Peschel, A. and Gotz, F. Analysis of the <i>Staphylococcus epidermidis</i> genes <i>epiF</i> , <i>-E</i> , and <i>-G</i> involved in epidermal immunity <i>J. Bacteriol.</i> 178 (2), 531-536 (1996)	2 (bases 2473 to 5292) Peschel, A., Schnell, N., Hille, M., Entian, K.D. and Gotz, F. Sequence of the <i>lan</i> antibiotics, epidermin and gallidermin: sequence determination of the <i>lan</i> operon

JOURNAL	production and their regulation by EpiO
MEDLINE	Mol. Gen. Genet. 254 (3), 312-318 (1997)
REFERENCE	97294510
AUTHORS	3 (bases 1 to 5297)
TITLE	Peschel, A., Schnell, N., Hille, M., Entlan, K.-D. and Goetz, F.
JOURNAL	Direct Submission
COMMENT	Submitted (15-JUN-1996) Mikrobielle Genetik, Universitaet Tuebingen, Waldhauser Strasse 70/8, Tuebingen 72076, Germany On Jun 14, 1997 this sequence version replaced gi:1176398. Other Staphylococcus epidermidis plasmid pue32 lantibiotic epidermin genes can be found in sequence with Genbank Accession Number X62386.

FEATURES	Location/Qualifiers
source	1, .5292

gene	complement(313..1005)
gene	/gene="epig"
CDS	complement(313..1005)
	/gene="epig"
	/function="Involved in conferring immunity towards the
	antibiotic epidermin"
	/note="Epig; contains integral membrane domain"
	/codon_start=1
	/transl_table=1
	/product="putative ABC transporter subunit"
	/protein_id="ABG164.1"
	/db_xref="PID:g1176401"
	/db_xref="GI:1176401"
	/translation="MKNELKVINLKRREILTYPTLLTLYIIFINLYLKNKNNTS
	ISQYSSNMAISBILATVYQVHPFESIGHFNHILGKPRINLVATLYISNLT
	FCILSTINFLMSQNLTLFVYISSFNMIILILFISLETGKISVAGSVVLT
	IFNIFGIEVLGDKSWYIPITVSTRYSMPINNSVPFLTISIVSVIIFLSLL
	LVKKSGRVD"
gene	complement(995..1759)
CDS	/gene="epib"
	complement(995..1759)

```

/function="involved in conferring immunity towards the
lactibiotic epidermin"
/note="EPIE: contains integral membrane domain"
/codon_start=1
/transl_table=11
/product="putative ABC transporter subunit"
/protein_id="AA61162.1"
/db_xref="PIR:G1176400"

```

```

/db_xref="GI:1176400"
/translation="KREIKAEVLRKFKNTTAMVFLSPFLGFLPALRQSPNK
DIDISFLLEHMPYFVWPIITICACSSMIGIERNGLFNYSNNMSINKMRSKI
LIMSJALLHLMELVALTGMMVGSGMSLILFLFLVFLSLPLILFLV
CGFVLSIFLIALTLISIIIFITMSKFWIMFWSYIGRIPLITLSLNGTMSKSS
YENDNSALATIVISIIYESIFEMSNNKSNMINXK"
complement(1761..2456)
/gene="epif"
complement(1761..2456)
/gene="epif"
/function="involved in conferring immunity towards the
antibiotic epidermin"
/feature="EpiF: contains ATP-binding domain"
/codon_start=1
/transl_table=11
/product="putative ABC transporter subunit"
/protein_id="AAB61163.1"
/db_xref="PID:g1176399"
/db_xref="GI:1176399"
/translation="MTYKIKETRNLRNKKFKETIIKDKINLKYRNEYVGLGINGACK
STLKITIGILPOTSGLDYLDSPLTRKELTIGSLSPPTYNLSAQDLKVALN
ENIDNEISKVELNVLNVEPQKVKNSLQKQRLGAMALIKPKLLVDESNGL
DPYGIQELRELKSLTNTNTSYIISHSLSIQOLADHYGIIHNGKLEYOEKNTDEN
LEDVFNITKSGK"
2615..3607
/gene="epih"
2615..3607
/gene="epih"
/function="involved in epidermin secretion"
/feature="EpiH"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAB61159.1"
/db_xref="PID:g2196513"
/db_xref="GI:2196513"
/translation="MYFLPLFVPLVITAIILFKRHSRLNTIYATQIVGLCFYV
ILAVFNSIVILMTWILSLILFLFNSKRLKREGSKLSNLSGLSILL
WIYTSPLSFINKSLAVIYVIFITVSLTFLSFLSCFISLFLSRKIDMSIS
YIVAGSLIEDKVPPLQSRIDIDKKNWDSKDKHTFLLSGGQKDESEQAMK
KYLIAQVNEKRIICENKSOSEYENLAKFSKLSIKNNATIIICSNYHILRAVFLAKL
KYSHFYSLGSKNKFENAFIRELVAILKMDWIFLCIVILLFWMIFTLIK"
complement(3744..4751)
/gene="epit"
complement(3744..4751)
/gene="epit"
/function="involved in epidermin secretion"
/feature="3' end of the epif gene which is disrupted by a
frame shift mutation"
/codon_start=1
/transl_table=11
/product="EpiT"
/protein_id="AAB61160.1"
/db_xref="PID:g2196514"
/db_xref="GI:2196514"
/translation="MTLVCTPLIIVILPIGINMKRVSKEGAPAKISYYNSR
LSIKLITLSTYNIKIKNTLAKNIDIDELHAKRYLSEFPMINILFLPNTIGILE
LGIYLMENMKSGDMFAIVLYLQIINPIVISTISYEVQRAIGSSRLIKKEPE
ELVITKITYNNVOGMEINDNTFTKDNKQIINSISLDAKQIYINIGESGCTEL
NILAGLNTGYTGNILDKRSQFSKYMRLFSYIQDLQILDVYNNIYINEN
ISIEIIONACKKTNLFTIOLNKSFSISIPDSINISIGOKRLVLRFLQKNLLE
Y"
complement(4759..5109)
/gene="epit"
complement(4759..5109)
/gene="epit"
/function="involved in epidermin secretion"
/feature="5' end of the epif gene which is disrupted by a
frame shift mutation"
/codon_start=1
/transl_table=11
/product="EpiT"
/protein_id="AAB61161.1"

```

```

/db_xref="PID:g2196515"
/db_xref="GI:2196515"
/translation="MIGTSFLYIPLIIRNALNKSLSIDKIVIIIIICFGLTLFSGV
STYLGIQGIQIIONISVWNNKIKPLSEFHLKNSASNLTSRLVNDTMITRFSVE
FIFSYSTNFIYV"
BASE COUNT 1944 a 645 c 658 g 2045 t
ORIGIN

```

```

Query Match 80.9%, Score 18.6, DB 1; Length 5292;
Best Local Similarity 76.2%, Pred. No. 25;
Matches 16; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
OY 1 AARATGAGTGGATTTTCAY 21
||:||||:||||:||||:
Db 3530 AAAATGATGGATTTTCAT 3550

```

```

RESULT 10
HS69B13
LOCUS HS69B13 310617 bp DNA HTG 10-JUN-1999
DEFINITION Homo sapiens chromosome 6 clone 69B13, WORKING DRAFT SEQUENCE, in
unordered pieces.
ACCESSION AL035698
NTD 95050992
VERSION AL035698.6 GI:5050992
KEYWORDS HTG; HTGS_PHRASEL.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 310617)
REFERENCE
AUTHORS Bates,K.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humangry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 11, 1999 this sequence version replaced gi:4902683.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n s separate
segments. Unfinished: dJ69B13 Contig_ID: 03436 acc-AL035698
Length: 304718 bp Unfinished: dJ69B13 Contig_ID: 02949
acc-AL035698 Length: 2867 bp Unfinished: dJ69B13 Contig_ID: 00022
acc-AL035698 Length: 1432 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
LOCATION/Qualifiers
1..310617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="69B13"
BASE COUNT 92374 a 57041 c 59286 g 100315 t 1601 others
ORIGIN

```

```

Query Match 78.3%; Score 18; DB 34; Length 310617;
Best Local Similarity 72.7%; Pred. No. 81;
Matches 16; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
OY 1 AARATGAGTGGATTTTCAY 22
||:||||:||||:||||:
Db 268920 AAAATGAGTGGATTTTCACA 268941

```

```

RESULT 11
DMU12634

```



```

repeat_region /map="7921-q22"
563..794
/rpt_family="L1"
repeat_region /rpt_family="L1"
1243..1272
/rpt_family="L1"
repeat_region /rpt_family="L1"
6077..6102
/rpt_family="L1"
repeat_region /rpt_family="L1"
13493..13826
/rpt_family="L1"
repeat_region /rpt_family="L1"
15133..15424
/rpt_family="ALU"
repeat_region /rpt_family="L1"
20066..20485
/rpt_family="L1"
repeat_region /rpt_family="L1"
25026..25305
/rpt_family="ALU"
misc_feature /rpt_family="ALU"
29019..29203
/note="match to EST D20877 (NID:9504697)"
repeat_region /rpt_family="L1"
36139..36166
/note="match to EST D20858 (NID:9504678)"
repeat_region /rpt_family="L1"
37435..37522
/rpt_family="ALU"
repeat_region /rpt_family="ALU"
37570..37667
/rpt_family="ALU"
repeat_region /rpt_family="ALU"
37670..37708
/rpt_family="ALU"
repeat_region /rpt_family="ALU"
38413..38687
/rpt_family="ALU"
repeat_region /rpt_family="ALU"
40439..40727
/rpt_family="ALU"
repeat_region /rpt_family="ALU"
42843..42878
/rpt_family="L1"
misc_feature /rpt_family="L1"
43587..43791
/note="match to EST D20858 (NID:9504678)"
repeat_region /rpt_family="PTR"
45232..45303
complement(50142..50216)
/rpt_family="PTR"
repeat_region /rpt_family="L1"
51297
complement(60928..61221)
/rpt_family="ALU"
repeat_region /rpt_family="L1"
62463..62800
/rpt_family="L1"
repeat_region /rpt_family="L1"
65195..66897
complement(65195..66897)
/rpt_family="PTR"
repeat_region /rpt_family="PTR"
65222..66897
/rpt_family="PTR"
repeat_region /rpt_family="L1"
71038..71064
/rpt_family="L1"
repeat_region /rpt_family="L1"
75188..75478
complement(81096..81385)
/rpt_family="ALU"
repeat_region /rpt_family="L1"
82157..82212
/rpt_family="L1"
repeat_region /rpt_family="L1"
92199..92490
/rpt_family="ALU"
repeat_region /rpt_family="L1"
95111..95142
/rpt_family="L1"
repeat_region /rpt_family="L1"
95520..95820
/rpt_family="ALU"
repeat_region /rpt_family="L1"
96048..96393
/rpt_family="MER"
repeat_region /rpt_family="MER"
96861..96929
/rpt_family="ALU"
repeat_region /rpt_family="ALU"
96932..96977
/rpt_family="L1"
repeat_region /rpt_family="L1"
102621..102621
complement(104717..104998)
/rpt_family="ALU"

```

```

BASE COUNT      34576 a      20570 c      20710 g      33022 t
ORIGIN
Query Match
Best Local Similarity 75.7%; Score 17.4; DB 11; Length 108878;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY      1 AARATGGATGATHTTTCATAC 23
Db 56039 AARATGGATGATTCATAC 56017
|||||:|:|:|:|
RESULT 15
LOCUS HUAC004605.259474 bp DNA PRI 24-JUL-1998
DEFINITION Homo sapiens Chromosome 16 BAC clone C19875K-A-248F7, complete
sequence.
ACCESSION AC004605
VERSION 93337395
KEYWORDS AC004605.1 GI:3337395
SOURCE HTG.
ORGANISM human.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 259474)
Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
Homo sapiens Chromosome 16 BAC clone C19875K-A-248F7
Unpublished
2 (bases 1 to 259474)
Adams,M.D. and Loftus,B.J.
Direct Submission
Submitted (23-APR-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, Email:
b.loftus@tigr.org
3 (bases 1 to 259474)
Adams,M.D. and Loftus,B.J.
Direct Submission
Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jul 24, 1998 this sequence version replaced gi:3228501.
Address all correspondence to: Mark Adams The Institute for Genomic
Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
address: humgent@tigr.org. The orientation of the sequence is from
SP6 end to T7 end. Genes were identified by a combination of five
methods including: XGRATL (available by anonymous ftp from
arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), GenScan (Chris Burge,
http://genomic.stanford.edu/~chris/GENSCANW.html) searches of the
complete sequence against a peptide database, and the Human gene
index database at TIGR (http://www.tigr.org/tib/tgi.html).
Genes without peptide homology having spliced EST hits are termed
'Unknown gene product'. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES
Source
1..259474
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q22-q23"
/clone="A-248F7"
17061..17175
/note="32160, SHGC-18227, Chr. 16, Homo sapiens"
/db_xref="dbSTS:G19282"
46411..46528
/note="7608, STS1-csRL-24g1-uA/csRL-24g1-uZ, Chr. -, Homo
sapiens"
/db_xref="dbSTS:G02122"
71868..71999
/note="32826, A002D07, Chr. -, Homo sapiens"
/db_xref="dbSTS:G19948"
STS

```

STS 72589. 72759
 /note="42758, D1S1053, Chr. -, Homo sapiens"
 /db_xref="dbSTS:628812"
 BASE COUNT 74059 a 49679 c 51651 g 84031 t 54 others
 ORIGIN

Query Match 75.7%; Score 17.4; DB 11; Length 259474;
 Best Local Similarity 69.6%; Pred. No. 1.6e+02;
 Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AARATGGATGATTTTCAYAC 23
 ||: ||| :||||: ||: ||
 Db 226583 AAGTGGGCTGGATATTTCACAC 226561

Search completed: September 13, 1999, 15:55:13
 Job time: 4541 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 14:39:33 ; Search time 255.05 Seconds

(without alignments)
22.562 Million cell updates/sec

Title: US-09-325-095-12

Perfect score: 23

Sequence: 1 AARATGATGATGATHTTCATC 23

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.6	89.6	3412	1	OS1156 Human p110 cDNA. R
2	20.6	89.6	3207	1	OS1155 p110 cDNA. Recombi
3	20.6	89.6	23	1	OS5012 Primer for p110 CD
4	20.6	89.6	3498	1	OS5012 Ptdins 3-Kinase 11
5	20.6	89.6	23	1	OS5018 Ptdins 3-Kinase pr
6	16.4	71.3	40875	1	TE80043 Insert from cosmid
7	16.4	71.3	381	1	TE67697 H. pylori secreted
8	16.4	69.6	516	1	TE67862 H. pylori secreted
9	16.4	69.6	800	1	TE78769 MART-1 antigenic p
10	15.8	68.7	339	1	TE78769 H. pylori GHPO 669
11	15.8	68.7	400	1	TE78769 Staphylococcus aur
12	15.8	68.7	1959	1	TE78769 Enterococcus faeca
13	15.8	68.7	5855	1	TE78769 Enterococcus faeca
14	15.8	68.7	48	1	TE78769 BclI methylase gen
15	15.6	67.8	3969	1	TE78762 Chicken protein ty
16	15.6	67.8	3969	1	TE78762 Chicken transmembr
17	15.6	67.8	110000	1	TE78762 Continuation (3 of
18	15.4	67.0	5113	1	TE78762 Human tyrosine dep
19	15.4	67.0	1701	1	TE78762 Mouse p53 binding
20	15.4	67.0	4752	1	TE78762 Nucleotide sequenc
21	15.4	67.0	4815	1	TE78762 Nucleotide sequenc
22	15.4	67.0	4799	1	TE78762 Nucleotide sequenc
23	15.4	67.0	1497	1	TE78762 H. pylori cell env
24	15.4	67.0	1587	1	TE78762 H. pylori cell env
25	15.4	67.0	1513	1	TE78762 Streptococcus pneu
26	15.4	67.0	6186	1	TE78762 Streptococcus pneu
27	15.4	67.0	1524	1	TE78762 DNA encoding a S.
28	15.4	67.0	874	1	TE78762 Clathrin-associat
29	15.4	67.0	6835	1	TE78762 Enterococcus faeca
30	15.4	67.0	7947	1	TE78762 Enterococcus faeca
31	15.4	67.0	2111	1	TE78762 H. pylori GHPO 103
32	15.4	67.0	1687	1	TE78762 H. pylori GHPO 170
33	15.4	67.0	3252	1	TE78762 Maize phosphatidyl
34	15.2	66.1	4157	1	TE78762 Maize-2 gene. Nucle
35	15.2	66.1	1640	1	TE78762 Maize-3 cDNA. Nucle
36	15.2	66.1	2226	1	TE78762 Maize-5 cDNA. Nucle
37	15.2	66.1	2305	1	TE78762 Maize-5 genomic DN
38	15.2	66.1	1844	1	TE78762 Parainfluenza viru
39	15.2	66.1	4157	1	TE78762 Tumour rejection a
40	15.2	66.1	1640	1	TE78762 Tumour rejection a
41	15.2	66.1	2226	1	TE78762 Tumour rejection a
42	15.2	66.1	2305	1	TE78762 Tumour rejection a
43	15.2	66.1	998	1	TE78762 Joined p1cr2, whic

ALIGNMENTS

C	44	15.2	66.1	2305	1	T01165
C	45	15.2	66.1	19932	1	T46159
						MAGE-51 gene. Dete
						CagI locus. Helico
RESULT	1					
OS1156						
ID	OS1156 standard; cDNA; 3412 BP.					
AC	OS1156;					
DT	12-APR-1994 (first entry)					
DE	Human p110 cDNA.					
KW	Phosphoinositide kinase; PI; p85 subunit; screening; agonist;					
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;					
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;					
KW	ds.					
OS	Human.					
PH	Key					
FT	Location/Qualifiers					
FT	1..3207					
FT	/*tag= a					
FT	/note= "PI3- kinase p110"					
DR	W09321328-A.					
PT	28-OCT-1993.					
PR	13-APR-1993; G00761.					
PR	13-APR-1992; GB-008135.					
PA	(LUDW-) LODWIG INST CANCER RES.					
PI	Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;					
PI	Parker PJ, Volinia S, Waterfield MD;					
DR	WPI: 93-351738/44.					
DR	P-PSDB: R43342.					
FT	Recombinant polypeptide(s) - with phosphoinositide-3 kinase					
FT	activity, useful for controlling cell proliferation					
PS	Claim 7; Fig 16; 146pp; English.					
CC	Southern blot analysis was performed using a bovine cDNA probe conty.					
CC	a fragment of a PI3-kinase-encoding sequence and human cDNA isolated					
CC	from a cDNA library constructed from mRNA isolated from the human					
CC	cell line K562. Positive clones were sequenced to give the human					
CC	PI3 kinase p110 sequence shown. This sequence has 95 percent					
CC	homology with the bovine sequence. The domain encoding residues 19-					
CC	100 of human p110 is sufficient to encode the kinase which will					
CC	associate with the p85 kinase subunit. The gene may be used to					
CC	provide a protein with PI3 kinase activity and is useful for					
CC	screening for (ant)agonists of PI3 kinase activity which could be					
CC	useful for stimulation or inhibition of cell proliferation and hence					
CC	prophylaxis or therapy. Platelet or neutrophil activity or blood					
CC	glucose levels can be controlled using the kinase.					
CC	See also OS1155 and OS7522-3.					
CC	Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;					
SQ						
Query Match	89.6%; Score 20.6; DB 1; Length 3412;					
Best Local Similarity	78.3%; Pred. No. 0.32;					
Matches	18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;					
QY	1 AARATGATGATGATHTTCATC 23					
DB	3160 AARATGATGATGATHTTCATC 3182					
RESULT	2					
OS1155						
ID	OS1155 standard; cDNA; 3207 BP.					
AC	OS1155;					
DT	12-APR-1994 (first entry)					
DE	p110 cDNA.					
KW	Phosphoinositide kinase; PI; p85 subunit; screening; agonist;					
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;					
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;					
KW	ss.					
PH	Key					
FT	Location/Qualifiers					
FT	1..3207					
FT	/*tag= a					

PR	05-AUG-1992; GB-016654.
PA	(IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PI	Goode NT, Nurse PM, Parker PJ, Waterfield MD;
PP	1991; 94-065697/08.
PT	Eukaryotic cells transformed with mammalian phospholipid or
PT	protein kinase DNA - useful in assays for compounds involved in
PT	cell growth regulation and for treating cancers
PS	Example 1; Page 14; 71pp; English.
CC	The sequences given in Q57018-19 are probes which were used in the
CC	isolation of fusion sequences comprising the phosphatidyl inositol
CC	(Pisins) 3-kinase under the regulatory control of the nmt promoter.
CC	These sequences were transformed into Schizosaccharomyces pombe cells
CC	in an embodiment of the invention. In the presence of thiamine the
CC	promoter is inactive and the cells carrying the pPC plasmids grow as
CC	the parental strain. In the absence of thiamine the nmt promoter
CC	functions and the pPC is induced. pPC activity is substantially
CC	increased under these conditions. Cells containing constructs such as
CC	this, are useful in assays for detecting compounds involved in cell
CC	growth regulation. It is also used as the basis for detecting
CC	compounds for treating cancers and the formation of blood vessel
CC	plaques.
SO	Sequence 23 BP; 7 A; 2 C; 4 G; 5 T;
QY	Query Match 89.6%; Score 20.5; DB 1; Length 23;
	Best Local Similarity 100.0%; Pred. No. 0.14;
	Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB	1 AARATGAGATGGATTTTCATC 23
	1 AARATGAGATGGATTTTCATC 23
RESULT	6
ID	T80043
AC	T80043 standard; DNA; 40875 BP.
DT	03-NOV-1997 (first entry)
DE	Insert*from cosmid 109.
KM	Pectinase; yeast; polygalacturonase; phytopathogen; vegetable spoilage;
KW	fruit spoilage; food processing; fruit juice; cosmid; ss.
OS	Saccharomyces cerevisiae.
FH	Key Location/Qualifiers
FT	cds 312..1943
FT	/*tag- a
FT	/note- "indicated in the specification as an open reading
FT	frame, encoding a protein with homology to
FT	S. cerevisiae allantoate permease (DML5)"
FT	cds 3622..7107
FT	/*tag- b
FT	/note- "indicated in the specification as an open reading
FT	frame"
FT	cds 3965..5188
FT	/*tag- c
FT	/note- "indicated in the specification as an open reading
FT	frame"
FT	cds 3461..4546
FT	/*tag- d
FT	/note- "indicated in the specification as an open reading
FT	frame, encoding a protein with homologies to
FT	pectinases"
FT	cds 5151..6194
FT	/*tag- e
FT	/note- "indicated in the specification as an open reading
FT	frame, with homology to S. pombe thiamine
FT	repressed gene"
FT	cds 5486..5845
FT	/*tag- f
FT	/note- "indicated in the specification as an open reading
FT	frame"
FT	cds 6386..7414
FT	/*tag- g
FT	/note- "indicated in the specification as an open reading

FT	frame"	.	7470	
FT	cds	6430.	7470	
FT		/tag	h	
FT		/note	"indicated in the specification as an open reading	
FT		frame"		
FT	cds	8051.	.8917	
FT		/tag	i	
FT		/note	"indicated in the specification as an open reading	
FT	cds	9656.	.10552	
FT		/tag	j	
FT		/note	"indicated in the specification as an open reading	
FT		frame"		
FT		/tag	k	
FT		/note	"indicated in the specification as an open reading	
FT	cds	10127.	.10936	
FT		/tag	l	
FT		/note	"indicated in the specification as an open reading	
FT		frame"	encoding a protein with homology to the	
FT		s. cerevisiae mitochondrial genome maintenance		
FT		protein (MGMI01)"		
FT	cds	11161.	.11523	
FT		/tag	m	
FT		/note	"indicated in the specification as an open reading	
FT		frame"		
FT	cds	13441.	.14517	
FT		/tag	n	
FT		/note	"indicated in the specification as an open reading	
FT	cds	14058.	.14360	
FT		/tag	o	
FT		/note	"indicated in the specification as an open reading	
FT	cds	14989.	.16119	
FT		/tag	p	
FT		/note	"indicated in the specification as an open reading	
FT		frame"		
FT	cds	16405.	.17619	
FT		/tag	q	
FT		/note	"indicated in the specification as an open reading	
FT		frame"		
FT	cds	16491.	.17336	
FT		/tag	r	
FT		/note	"indicated in the specification as an open reading	
FT		frame"	encoding a protein with homology to the	
FT		s. cerevisiae ribosomal protein S7A (SCRPS7A)"		
FT	cds	18429.	.18797	
FT		/tag	s	
FT		/note	"indicated in the specification as an open reading	
FT	cds	18749.	.21037	
FT		/tag	t	
FT		/note	"indicated in the specification as an open reading	
FT	cds	21763.	.22488	
FT		/tag	u	
FT		/note	"indicated in the specification as an open reading	
FT	cds	22540.	.23262	
FT		/tag	v	
FT		/note	"indicated in the specification as an open reading	
FT	cds	23667.	.28607	
FT		/tag	w	
FT		/note	"indicated in the specification as an open reading	
FT		frame"		
PN	W09704106-A2.			
PD	06-FEB-1997.			
PF	15-JUL-1996.			
PF	22-DEC-1995.			
PR	14-JUL-1995.			
PR	EP-870089.			
PR	EP-870089.			

PA (INNO-) INNOGENETICS NV..
 PI Scarce T, Van Broekhoven A;
 DR WPI: 97-132645/12.
 CC New nucleic acid encoding *Saccharomyces pectinase* - useful in food
 PT processing, also its regulatory sequences are useful for expression
 PT of heterologous protein
 PS Example 1: Fig 2: 65pp; English.
 CC This sequence represents the insert from cosmid 109, containing
 CC *Saccharomyces cerevisiae* strain Y14679 DNA. This sequence contained the
 CC pectinase gene of the invention (see 780042). Pectinases are synthesised
 CC by plants and a variety of microorganisms, the best characterised
 CC pectinase being polygalacturonase. Pectinases play roles in the invasion
 CC of plant tissues by phytopathogens, the spoilage of fruits and
 CC vegetables, and also in their food processing and plant biological
 CC applications. Vectors containing the pectinase coding sequence are used
 CC for production of homologous or heterologous proteins in transformed host
 CC cells. The pectinase sequence may include regulatory and/or coding
 CC sequences, including the sequence encoding the pectinase signal peptide
 CC (which can be used in any recombinant system). The pectinase is useful in
 CC food processing (e.g. preparation of fruit juices), typical heterologous
 CC proteins that can be made are (poly)peptides of bacterial, animal or
 CC viral origin. Also, probes derived from the pectinase coding sequence can
 CC be used to detect the coding sequence and primers for amplification are
 CC used to isolate related genes.
 SQ Sequence 40875 BP; 12953 A; 7579 C; 8292 G; 12051 T;

Query Match 71.3%; Score 16.4; DB 1; Length 40875;
 Best Local Similarity 68.2%; Pred. No. 59;
 Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 AARATGATGATGATHTTTCAYA 22
 DB 30251 AARATGACTGATCTTCATA 30272

RESULT 7
 T67697 standard; DNA: 381 BP.
 AC T67697;
 DT 15-JUL-1997 (first entry)
 DE H. pylori secreted or periplasmic protein ORF 5869090. aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
 OS *Helicobacter pylori*.
 FH Key Location/Qualifiers
 FT cds 1..381
 FT /*tag= a
 FT /note= "no stop codon given"
 PN W09640893-A1.
 PD 19-DEC-1996.
 PE 06-JUN-1996; 009122.
 PF 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR P-PSDB: W20547.
 PT *Helicobacter pylori* nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent *H. pylori*
 PT infection, and to detect *Helicobacter*
 PS Claim 23; Page 308; 1481pp; English.
 CC This sequence encodes a *H. pylori* secreted or periplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat *H. pylori*
 CC infection or to identify *H. pylori* polypeptide binding compounds.
 CC useful as potential *H. pylori* life cycle activators or inhibitors.
 CC The genomic sequence of *H. pylori* (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely *H. pylori* antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide
 CC production, e.g. in *E. coli* hosts.
 SQ Sequence 381 BP; 133 A; 59 C; 71 G; 118 T;

Query Match 69.6%; Score 16; DB 1; Length 381;
 Best Local Similarity 77.8%; Pred. No. 41;
 Matches 14; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 5 TGGATGATGATHTTTCAYA 22
 DB 60 TGGATGATGATTTTCACA 77

RESULT 8
 T67862 standard; DNA: 516 BP.
 AC T67862;
 DT 14-JUL-1997 (first entry)
 DE H. pylori secreted or periplasmic protein ORF 019p11016orf14.
 KW Vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
 KW secreted; periplasmic; ds.
 OS *Helicobacter pylori*.
 FH Key Location/Qualifiers
 FT cds 1..516
 FT /*tag= a
 FT /note= "no stop codon given"
 PN W09640893-A1.
 PD 19-DEC-1996.
 PE 06-JUN-1996; 009122.
 PF 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR P-PSDB: W20609.
 PT *Helicobacter pylori* nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent *H. pylori*
 PT infection, and to detect *Helicobacter*
 PS Claim 23; Pages 753-754; 1481pp; English.
 CC The present sequence encodes a *Helicobacter pylori* secreted or
 CC periplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat *H. pylori*
 CC infection or to identify *H. pylori* polypeptide binding compounds.
 CC useful as potential *H. pylori* life cycle activators or inhibitors.
 CC The genomic sequence of *H. pylori* (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the
 CC bacterial DNA. The sequences were analysed for ORF of at least 180
 CC nucleotides, and the predicted coding regions defined by computer
 CC evaluation. To identify likely *H. pylori* antigens for vaccine
 CC development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences
 CC of interest, particular regions can be isolated from *H. pylori* by
 CC PCR amplification for recombinant polypeptide production, e.g. in
 CC *E. coli* hosts.
 SQ Sequence 516 BP; 184 A; 69 C; 106 G; 157 T;

Query Match 69.6%; Score 16; DB 1; Length 516;
 Best Local Similarity 77.8%; Pred. No. 43;
 Matches 14; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 5 TGGATGATGATHTTTCAYA 22
 DB 63 TGGATGATGATTTTCACA 80

RESULT 9

T78769
ID T78769 standard; cDNA; 800 BP.
AC T78769;
DT 02-MAR-1998 (first entry)
DE MART-1 antigenic protein nucleic acid melanoma antigen.
KM MART-1 antigen; melanoma; dendritic cell; T cell; recombinant;
KW haematopoietic stem cell; antigenic peptide; cancer; MHC;
OS major histocompatibility complex; ss.
PN Mus musculus.
PD WO9729183-A2.
PF 14-AUG-1997.
PR 07-FEB-1997; U02063.
PS 08-FEB-1996; US-011433.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Hwu P, Reeves M, Rosenberg SA.
PT New transformed dendritic cells for activating T cells - produced by
PT transducing haematopoietic stem cells with nucleic acid and
PT differentiating the cells into dendritic cells
PS Disclosure; Fig 3; 63pp; English.
CC A novel method has been developed for making a dendritic cell transduced
CC with a selected nucleic acid. The method comprises: (a) transducing a
CC haematopoietic stem cell (HSC) with the selected nucleic acid; and (b)
CC differentiating the transduced stem cell into a dendritic cell, thereby
CC making a dendritic cell transduced with the selected nucleic acid. The
CC present sequence represents the nucleic acid sequence for MART-1 which
CC encodes an antigenic protein, a major histocompatibility complex class
CC I bound melanoma antigen. The recombinant dendritic cells expressing
CC antigenic peptides can activate T-cells against target cells expressing
CC selected antigens in vivo. They can also be used for detecting T cell
CC mediated anti-cancer cell activity of a target antigenic peptide and
CC for altering the MHC class II antigenic repertoire of a dendritic cell.
CC They can be used for the treatment of cancers and cellular infections
CC and in diagnostic and cell screening assays. T cells activated by
CC contact with dendritic cells may be used to kill a target cell,
CC especially a cancer cell. The recombinant dendritic cells provide
CC broader antigen expression and more efficient MHC class I and class II
CC peptide loading, and the ability to expand the population of desired
CC dendritic cells; e.g. in culture. 175 C; 205 G; 191 T;
SQ Sequence 800 BP; 185 A; 175 C; 205 G; 191 T;

Query Match 69.6%; Score 16; DB 1; Length 800;
Best Local Similarity 63.6%; Pred. No. 46;
Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

2 AGATGCAVTCGATHTTCAYAC 23
1:||||:||||:||||:||||:
Db 195 AGATGCAVTCGATHTTCCTAC 216

RESULT 10
X14176/c
X14176 standard; DNA; 339 BP.
AC X14176;
DT 31-MAR-1999 (first entry)
DE H. pylori GPO 669 gene.
KM GPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT CDS 1..339
FT CD5
PN WO9843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PS 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HOMA-) HUMAN GENOME SCI INC.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI; 98-542293/46.

DR P-BSDB; W98457.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 1; Page 804; 2054pp; English.
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 339 BP; 112 A; 80 C; 48 G; 99 T;

Query Match 69.6%; Score 16; DB 1; Length 339;
Best Local Similarity 77.8%; Pred. No. 40;
Matches 14; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 5 TCGAATGCAVTCGATHTTCAYAC 22
||||:||||:||||:||||:
Db 147 TCGAATGCAVTCGATHTTCAC 130

RESULT 11
V75319/c
V75319 standard; DNA; 400 BP.
AC V75319;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus confg SEQ ID #1008.
KM Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
OS Staphylococcus aureus.
PN EP-76519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HOMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 1791; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 400 BP; 145 A; 72 C; 55 G; 123 T;

Query Match 68.7%; Score 15.8; DB 1; Length 400;
Best Local Similarity 57.1%; Pred. No. 52;
Matches 12; Conservative 8; Mismatches 1; Indels 0; Gaps 0;

RESULT 15

ID V05762 Standard: cDNA: 3969 BP.
 AC V05762:
 DT 12-MAY-1998 (first entry)
 DE Chicken protein tyrosine phosphatase-lambda gene.
 KM Chicken: protein tyrosine phosphatase-lambda; brain; probe; primer; PCR;
 KW hybridisation; human CD45; amplification; alternative splicing product;
 OS Gallus domesticus.
 FH Key Location/Qualifiers
 FT CDS 91..3804
 TT /frag= a
 PT /product= "protein tyrosine phosphatase
 PN US5693486-A.
 PD 02-DEC-1997.
 PF 12-MAY-1994; 241853.
 PK 12-MAY-1994; US-241853.
 PA (UVR0) UNIV ROCKEFELLER.
 PI Fang KS, Hanafusa H;
 DR WPI; 98-031746/03.
 DR P-PSDB; W44729.
 DT Nucleic acid molecule encoding chicken protein tyrosine phosphatase
 PT - specifically transmembrane protein tyrosine phosphatase-lambda,
 PT useful to develop probes and primers
 PS Claim 2; Column 31-40; 51pp; English.
 CC This is the nucleotide sequence which encodes the novel chicken protein
 CC tyrosine phosphatase (PTP)-lambda. The sequence was isolated from a
 CC chicken brain cDNA library, using, as a probe, a fragment encoding the
 CC intracellular domain of the human CD45 sequence. The sequence was then
 CC amplified using primers V05763-V05766. The gene has a transcript of
 CC around 5.6 kb and is abundant in spleen and intestine. The sequence has
 CC been shown to contain 5 alternative splicing products which vary near
 CC the N-terminus of the encoded protein. Nucleic acid molecules,
 CC especially encoding residues 22-1237, 22-509, 510-531 or 532-1237 of the
 CC 1237 residue PTP-lambda protein, can be used as probes and primers to
 CC detect levels of phosphatase expression. This is useful for the
 CC diagnosis and treatment of diseases such as cancer or immune system
 CC functional disorders. The protein can also be used to isolate ligands
 CC of the PTP.
 SQ Sequence 3969 BP; 1317 A; 816 C; 847 G; 989 T;

Query Match

67.88; Score 15.6; DB 1; Length 3969;

Best Local Similarity 75.08; Pred. NO. 97;

Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 RATGATGATHTTYCAY 22

Db 3265 AATGATTGATGTTCCAGA 3284

Search completed: September 13, 1999, 15:59:29
 Job time: 4796 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 14:38:32 ; Search time 1694.61 Seconds
(without alignments)
26.772 Million cell updates/sec

Title: US-09-325-095-12

Perfect score: 23
Sequence: 1 AARATGATGATHTTCATAC 23

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20.6	89.6	292	23	D56232	D56232 HOM420G03B
2	20.6	89.6	264	24	N23534	N23534 YV91G08..r1
3	20.6	89.6	475	25	N44819	N44819 YV39A10..r1
4	20.6	89.6	405	30	AA237307	AA237307 mw96h11.r
5	20.6	89.6	331	33	AA445047	AA445047 v957h11.r
6	20.6	89.6	206	39	AA874470	AA874470 vx80f10.r
7	20.6	89.6	447	42	A1155136	A1155136 u459f06.r
8	20.6	89.6	544	46	A1411730	A1411730 EST240024
C 9	20.6	89.6	438	47	A1502906	A1502906 UI-R-C1-K
C 10	20.6	89.6	431	48	A1547814	A1547814 UI-R-C3-S
C 11	20.6	89.6	633	49	A1632172	A1632172 t859g07.x
12	20	87.0	453	45	A1324274	A1324274 mq72c02.y
13	19	82.6	433	31	AA298585	AA298585 EST114216
14	18	78.3	433	26	w76227	w76227 zd64c12..r1
15	17.2	74.8	695	45	A1354431	A1354431 gu17c11.x
16	17	73.9	410	23	R83858	R83858 ypb2e10.r1
17	17	73.9	593	30	AA212806	AA212806 mw84e10.r
18	17	73.9	530	30	AA239420	AA239420 mw98d08.r
19	17	73.9	394	35	AA575527	AA575527 vm03b08.r
20	17	73.9	441	42	A1156560	A1156560 ue43e06.r
21	16.4	71.3	406	22	R73756	R73756 y155f02..r1
22	16.4	71.3	465	22	T75851	T75851 10629.Lambda
23	16.4	71.3	485	25	N44970	N44970 YV35g10..r1
C 24	16.4	71.3	410	33	AA443680	AA443680 zw66c09.s
C 25	16.4	71.3	326	35	AA585856	AA585856 28505.Lam
C 26	16.4	71.3	294	36	AA652512	AA652512 ns72a01.s
C 27	16.4	71.3	516	41	A1052103	A1052103 oy30a04.x
28	16.4	71.3	433	41	AU016499	AU016499 AU016499
29	16.4	71.3	584	41	AU016839	AU016839 AU016839
C 30	16.4	71.3	405	43	A1239854	A1239854 gha3h09.x
31	16.4	71.3	902	44	A1323676	A1323676 mq21g08.x
32	16.4	71.3	288	46	A1413704	A1413704 me03h10.x
33	16.4	71.3	503	47	A1523627	A1523627 t99sc06.x
34	16.4	71.3	243	48	A1605101	A1605101 vho04d1.x
C 35	16.4	71.3	471	50	A1692399	A1692399 wd68b03.x
C 36	16.4	71.3	560	50	A1692482	A1692482 wd70h12.x
C 37	16.4	71.3	206	50	AV029255	AV029255 AV029255
38	16.2	70.4	438	29	AA127013	AA127013 z116h06.s
C 39	16.2	70.4	525	30	AA260539	AA260539 vbo0e07.r
40	16.2	70.4	430	33	AA427986	AA427986 zw50b01.r
C 41	16.2	70.4	349	37	AA727030	AA727030 vu36h05.r
C 42	16.2	70.4	439	43	A1187390	A1187390 qf31a02.x
C 43	16.2	70.4	474	47	A1644790	A1644790 mz81a05.y
C 44	16.2	70.4	436	47	A1485837	A1485837 EST244158
C 45	16	69.6	183	20	T27380	T27380 hbc2456.Hum

ALIGNMENTS

RESULT 1	D56232/c	292 bp	MRNA	EST	31-AUG-1995
LOCUS	D56232				
DEFINITION	HOM420G03B Clontech human fetal brain polyA+ mRNA (#6535)				Homo
ACCESSION	D56232				
NID	9970756				
VERSION	D56232.1	GI:970756			

NAME: EST.
 SOURCE: human.
 ORGANISM: Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE
 1 (bases 1 to 292)
 Fujiiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,
 Okuno, S., Ozaki, K., Shimizu, F., Shimizu, H.,
 Takachi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y.,
 Mekawa, H., Shin, S. and Nakamura, Y.
 Fujiiwara et al. (1995)
 TITLE
 JOURNAL
 Unpublished (1995)
 COMMENT
 On Sep 21, 1992 this sequence version replaced g1:276155.
 Contact: Tsutomu Fujiiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawanchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035
 Insert Length: 770 Std Error: 0.00
 High quality sequence stop: 331.
 Location/Qualifiers
 1..292
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="Clontech human fetal brain polyA+ mRNA
 (#5535)"
 BASE COUNT
 78 a 59 c 55 g 100 t
 ORIGIN
 Query Match 89.6%; Score 20.6; DB 23; Length 292;
 Best Local Similarity 78.3%; Pired. No. 7.5;
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 AARATGATGATGATTCACAC 23
 11:|||||:|||||:|||||
 Db 67 AAATGATGATTCACAC 45
 RESULT 2
 N23534 264 bp mRNA EST 28-DEC-1995
 LOCUS
 Y91908.F1 Soares melanocyte 2NDHM Homo sapiens CDNA clone
 IMAGE:250142 5' similar to SP:P100_BOVIN P32871
 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT ; mRNA sequence.
 ACCESSION
 N23534
 NID G1137684
 VERSION N23534.1 GI:1137684
 KEYWORDS
 EST.
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE
 1 (bases 1 to 264)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasaki, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 TITLE
 JOURNAL
 Unpublished (1995)
 COMMENT
 On Apr 14, 1993 this sequence version replaced g1:692929.
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence stops: 225

Source: IMAGE Consortium, LNL.
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1135 Std Error: 0.00
 Seq primer: M13Rpl
 High quality sequence stop: 225.
 Location/Qualifiers
 1..264
 /organism="Homo sapiens"
 /db_xref="GDB:386784"
 /db_xref="taxon:9606"
 /map="14"
 /clone_id="IMAGE:250142"
 /clone_lib="Soares melanocyte 2NDHM"
 /sex="Male"
 /issue_type="melanocyte"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT7r3D (Pharmacia) with a modified
 polylinker. Site 1: Not I. Site 2: Eco RI. 1st strand CDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTTACCAATCTGAAGTGGAGCGCCGAGTCTTTTCTTTTCTTTT 3'],
 double-stranded CDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT7r3 vector
 (Pharmacia). Library constructed by Bento Soares and
 M. Fatima Bonaldo. RNA from normal foreskin melanocytes
 (#3374) was kindly provided by Dr. Anthony P. Albino."
 BASE COUNT
 94 a 53 c 52 g 60 t
 ORIGIN
 Query Match 89.6%; Score 20.6; DB 24; Length 264;
 Best Local Similarity 78.3%; Pired. No. 7.5;
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 AARATGATGATGATTCACAC 23
 11:|||||:|||||:|||||
 Db 140 AAATGATGATTCACAC 162
 RESULT 3
 N44819 475 bp mRNA EST 13-FEB-1996
 LOCUS
 Y939410.F1 Soares melanocyte 2NDHM Homo sapiens CDNA clone
 IMAGE:273594 5' similar to SW:P11A_BOVIN P32871
 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
 mRNA sequence.
 ACCESSION
 N44819
 NID G1185985
 VERSION N44819.1 GI:1185985
 KEYWORDS
 EST.
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE
 1 (bases 1 to 475)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasaki, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 TITLE
 JOURNAL
 Unpublished (1995)
 COMMENT
 On May 8, 1995 this sequence version replaced g1:800198.
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: 77
High quality sequence stop: 402.

FEATURES

source

Location/Qualifiers

1..475

/organism="Homo sapiens"
/db_xref="GDB:3883236"
/db_xref="taxon:9606"
/clone_image="273594"
/clone_lib="Soares melanocyte 2NDHM"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCGACATTTTCTTTTCTTTT 3'];
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT
ORIGIN

170 a 90 c 85 g 128 t 2 others

Query Match 89.6%; Score 20.6; DB 25; Length 475;
Best Local Similarity 78.3%; Pred. No. 7.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 AARATGATGATTTTCAYAC 23

Db 201 AARATGATGATTTTCACAC 223

RESULT 4

LOCUS

AA237307 405 bp mRNA EST 03-MAR-1997
m96h11.r1 Soares mouse NMU Mus musculus cDNA clone IMAGE:678557 5'
similar to SW:P1LA.BOVIN P32871 PHOSPHATIDYLINOSITOL 3-KINASE
CATALYTIC SUBUNIT, ALPHA ISOFORM ; , mRNA sequence.

ACCESSION

AA237307
91861328
AA237307.1 GI:1861328

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)

On Sep 12, 1996 this sequence version replaced gi:1404918.

TITLE

JOURNAL

COMMENT

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:418301
Putative full length read
vector to vector length is
Seq primer: -28m13 rev2 ET from Amersham.

FEATURES

Location/Qualifiers

source

1..405

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="678597"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCGACATTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adapters
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT 141 a 79 c 90 g 95 t

ORIGIN

Query Match 89.6%; Score 20.6; DB 30; Length 405;
Best Local Similarity 78.3%; Pred. No. 7.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 AARATGATGATTTTCAYAC 23

Db 25 AARATGATGATTTTCACAC 47

RESULT 5

LOCUS

AA445047 331 bp mRNA EST 03-JUN-1997
v957h11.r1 Beddington mouse embryonic region Mus musculus cDNA
clone IMAGE:865509 5' similar to SW:P1LA.MOUSE P42337
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ; ,
mRNA sequence.

AA445047
92157730
AA445047.1 GI:2157730

EST.
VERSION
KEYWORDS
SOURCE

house mouse.
Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 331)
Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

The WashU-HMI Mouse EST Project
Unpublished (1996)

On May 5, 1995 this sequence version replaced gi:798358.

On May 5, 1995 this sequence version replaced gi:798358.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.

```

/clone_lib="Beddington mouse embryonic region"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="7.5dpc"
/lab_host="DH12s"
/Note="Organ: whole embryo; Vector: PCMV-SPORT; Site:1:
oligo dr. Gastrulating embryos were collected at 7.5dpc
from C57BL6 x DBA matings, excluding embryos that had
developed head folds and all extraembryonic tissues.
Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
Referenced in Development 121, 2479-2489 (1995)."

BASE COUNT      106 a      64 c      72 g      89 t
ORIGIN

Query Match
Best Local Similarity 89.6%; Score 20.6; DB 33; Length 331;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY      1 AARATGATGATGATTTTCATAC 23
      11:|||||:|||||:|||||:|||||:
Db      288 AAATGATGATGATCTCCACAC 310

RESULT 6
LOCUS      AAB74470      206 bp      mRNA      EST      19-MAR-1998
DEFINITION vx80f10.r1 Soares 2NDMT Mus musculus cDNA clone IMAGE:1281547 5',
ACCESSION  AAB74470
VERSION     92979159
KEYWORDS    AAB74470.1 GI:2979159
SOURCE      EST.
ORGANISM    house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 206)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
            On Jan 14, 1998 this sequence version replaced gi:1798041.

TITLE
JOURNAL
COMMENT

FEATURES
            source
            1..206
            Location/Qualifiers
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /map="4"
            /clone_lib="Soares 2NDMT"
            /sex="male"
            /tissue_type="Thymus"
            /dev_stage="4 weeks"
            /lab_host="DH10B"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified

```

```

polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15.
TGTTCACATCGAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGTGTGT
3'}; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Falima Bonaldo."

BASE COUNT      61 a      47 c      57 g      41 t
ORIGIN

Query Match
Best Local Similarity 89.6%; Score 20.6; DB 39; Length 206;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY      1 AARATGATGATGATTTTCATAC 23
      11:|||||:|||||:|||||:|||||:
Db      3 AAATGATGATGATCTCCACAC 25

RESULT 7
LOCUS      A1155136      447 bp      mRNA      EST      30-SEP-1998
DEFINITION ud59f06.r1 Soares mouse uterus NMPu Mus musculus cDNA clone
IMAGE:1450211 5' similar to SW:P11A.MOUSE P42337
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;,
ACCESSION  A1155136
VERSION     93683605
KEYWORDS    A1155136.1 GI:3683605
SOURCE      EST.
ORGANISM    house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 447)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
            On Apr 18, 1995 this sequence version replaced gi:775514.

TITLE
JOURNAL
COMMENT

FEATURES
            source
            1..447
            Location/Qualifiers
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone_lib="Soares mouse uterus NMPu"
            /sex="female"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Organ: uterus; Vector: pT7T3D-Pac (Pharmacia) with
            a modified polylinker; 1st strand cDNA was prepared from
            pregnant mouse uterus, and was then primed with a Not I -
            oligo(dT) primer. Double-stranded cDNA was ligated to Eco

```

RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 139 a 79 c 101 g 127 t 1 others
 Query Match 89.6%; Score 20.6; DB 42; Length 447;
 Best Local Similarity 78.3%; Pred. No. 7.3;
 Matches 18; Conservative 5; Mismatches 0; Gaps 0;
 Indels 0;
 OY 1 AARATGATGATGATTCATCAYAC 23
 Db 403 AARATGATGATGATTCATCAYAC 425

RESULT 8
 LOCUS A1411730 544 bp mRNA EST 09-FEB-1999
 DEFINITION EST240024 Normalized rat placenta, Bento Soares Rattus sp. CDNA
 ACCESSION A1411730
 MID 94255234
 VERSION A1411730.1 GI:4255234
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 544)
 Lee, N.H., Glodet, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Keilavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 On Apr 7, 1998 this sequence version replaced gi:3035706.
 Other ESTs: EST240023
 CONTACT: Lee, NH
 ATCC

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source
 1..544
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RPDJ40"
 /clone_lib="Normalized rat placenta, Bento Soares"
 /note="Organ: Placenta; Vector: pRT3pac; Site_1: EcoRI;
 Site_2: NotI"
 Location/Qualifiers
 131 a 119 c 110 g 184 t

BASE COUNT 131 a 119 c 110 g 184 t
 ORIGIN
 Query Match 89.6%; Score 20.6; DB 46; Length 544;
 Best Local Similarity 78.3%; Pred. No. 7.2;
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Indels 0;
 OY 1 AARATGATGATGATTCATCAYAC 23
 Db 385 AARATGATGATGATTCATCAYAC 363

RESULT 9
 LOCUS A1502906 438 bp mRNA EST 11-MAR-1999
 DEFINITION UI-R-C1-Kn-b-09-0-UI.s1 UI-R-C1 Rattus norvegicus CDNA clone
 UI-R-C1-Kn-b-09-0-UI 3', mRNA sequence.
 ACCESSION A1502906

NTD 94400757
 VERSION A1502906.1 GI:4400757
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 438)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL
 MEDLINE
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 On Mar 16, 1998 this sequence version replaced gi:2961816.

Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. CDNA Library Preparation: M.B. Soares Lab Clone distribution:
 Clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..438
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10115"
 /map="7p14-15"
 /clone="UI-R-C1-Kn-b-09-0-UI"
 /clone_lib="UI-R-C1"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
 library is a subtracted library derived from the UI-R-C0
 library, which is a subtracted library derived from the
 UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
 consisted of a mixture of individually tagged normalized
 libraries constructed from rat placenta, adult lung,
 brain, liver, kidney, heart, spleen, ovary, and muscle.
 The UI-R-E1 library consisted of a mixture of
 individually tagged normalized libraries constructed from
 8, 12 and 18-day embryo. The tag is a string of 3-5
 nucleotides present between the Not I site and the
 oligo-dT track which allows identification of the library
 of origin of a clone within the mixture. The subtracted
 library (UI-R-C1) was constructed as follows: PCR
 amplified cDNA inserts from UI-R-C0 clones from which 3'
 ESTs had been derived was used as a driver in a
 hybridization with the UI-R-C0 library in the form of
 single-stranded circles. The remaining single-stranded
 circles (subtracted library) was purified by
 hydroxyapatite column chromatography, converted to
 double-stranded circles and electroporated into DH10B
 bacteria (Life Technologies) to generate the UI-R-C1
 library. This procedure has been previously described
 (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
 1996)."

FEATURES
 source
 1..438
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10115"
 /map="7p14-15"
 /clone="UI-R-C1-Kn-b-09-0-UI"
 /clone_lib="UI-R-C1"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
 library is a subtracted library derived from the UI-R-C0
 library, which is a subtracted library derived from the
 UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
 consisted of a mixture of individually tagged normalized
 libraries constructed from rat placenta, adult lung,
 brain, liver, kidney, heart, spleen, ovary, and muscle.
 The UI-R-E1 library consisted of a mixture of
 individually tagged normalized libraries constructed from
 8, 12 and 18-day embryo. The tag is a string of 3-5
 nucleotides present between the Not I site and the
 oligo-dT track which allows identification of the library
 of origin of a clone within the mixture. The subtracted
 library (UI-R-C1) was constructed as follows: PCR
 amplified cDNA inserts from UI-R-C0 clones from which 3'
 ESTs had been derived was used as a driver in a
 hybridization with the UI-R-C0 library in the form of
 single-stranded circles. The remaining single-stranded
 circles (subtracted library) was purified by
 hydroxyapatite column chromatography, converted to
 double-stranded circles and electroporated into DH10B
 bacteria (Life Technologies) to generate the UI-R-C1
 library. This procedure has been previously described
 (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
 1996)."

BASE COUNT 94 a 92 c 85 g 167 t
 ORIGIN
 Query Match 89.6%; Score 20.6; DB 47; Length 438;
 Best Local Similarity 78.3%; Pred. No. 7.3;
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Indels 0;
 OY 1 AARATGATGATGATTCATCAYAC 23

Db 399 AARATGACTGATCTTCACAC 377

RESULT 10
A1547814 431 bp MRNA EST 22-MAR-1999
DEFINITION
UI-R-C3-sk-b-10-0-01.g1 UI-R-C3 Rattus norvegicus cDNA clone
A1547814
NID 94465302
VERSION A1547814.1 GI:4465302
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 431)
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL
MEDLINE
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187473.

FEATURES
source
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
*tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized placenta library cDNA library preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com) The following repetitive elements were
found in this cDNA sequence: 1-70, >POLY_A#simple.repeat
Seq primer: M13 Forward.
Location/Qualifiers
1..431
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/map="1: 874H02; 7: 7421.3-7422.1; 4p16.3; 2"
/clone="UI-R-C3-sk-b-10-0-01"
/clone_lib="UI-R-C3"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C3
library is a subtracted library of a series, ultimately
derived from a mixture of individually tagged normalized
libraries from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, and 8, 12 and 18-day
embryos, after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: UI-R-C3, UI-R-C2p,
UI-R-C1, UI-R-C0, UI-R-A1, UI-R-E1. The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C3) was constructed as follows: PCR-amplified
cDNA inserts from UI-R-C2p clones from which 3' ESTs had
been derived was used as a driver in a hybridization with
the UI-R-C2p library in the form of single-stranded
circles. The remaining single-stranded circles (subtracted
library) was purified by hydroxyapatite column
chromatography, converted to double-stranded circles and

electroporated into DH10B bacteria (Life Technologies) to
generate the UI-R-C3 library. This procedure has been
previously described (Bonaldio, Lennon and Soares, Genome
Research 6:791-806, 1996)."

BASE COUNT 93 a 92 c 84 g 162 t
ORIGIN
Query Match 89.6%; Score 20.6; DB 48; Length 431;
Best Local Similarity 78.3%; Pred. No. 7.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
QY 1 AARATGACTGATCTTCACAC 23
Db 401 AARATGACTGATCTTCACAC 379

RESULT 11
A1632172 633 bp MRNA EST 26-APR-1999
LOCUS A1632172/c
DEFINITION
t885g07.x1 NCI-CGAP_GC6 Homo sapiens cDNA clone IMAGE:228108 3'
similar to SW:P1LA.BOVIN P32871 PHOSPHATIDYLINOSITOL 3-KINASE
CATALYTIC SUBUNIT, ALPHA ISOFORM ;, mRNA sequence.
A1632172
NID 94683502
VERSION A1632172.1 GI:4683502
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 633)
REFERENCE
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT On May 7, 1998 this sequence version replaced gi:3121411.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldio, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILM at:
www.bio.lnlnl.gov/bdrp/image/image.html

FEATURES
source
Seq primer: -400P from Gibco
High quality sequence stop: 446.
Location/Qualifiers
1..633
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="15q26.1"
/clone="IMAGE:228108"
/clone_lib="NCI-CGAP_GC6"
/tissue_type="Pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Plasmid DNA from the normalized library
NCI-CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (clones 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldio.

BASE COUNT 169 a 109 c 108 g 246 t 1 others
ORIGIN

Query Match 89.6%; Score 20.6; DB 49; Length 633;
 Best Local Similarity 78.3%; Pred. No. 7.1;
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 AARATGATGGATCTTCACAC 23
 Db 391 AARATGATGGATCTTCACAC 369

RESULT 12
 LOCUS A1324274 453 bp mRNA EST 23-DEC-1998
 DEFINITION mg72c02.y1 Stratiagene mouse melanoma (#937312) Mus musculus cDNA
 clone IMAGE:584258 5' similar to SW:PIA_MOUSE P42337
 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;,
 mRNA sequence.

ACCESSION A1324274
 MID g4058703
 VERSION A1324274.1 GI:4058703
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 453)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Gessel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Thibault,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 On Jan 14, 1998 this sequence version replaced gi:1798536.

TITLE JOURNAL
 COMMENT

Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:358906
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 Seq primer: -40RP from Gibco
 High quality sequence stop: 403.

FEATURES
 source 1..453
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="21"
 /clone="IMAGE:584258"
 /clone_lib="Stratiagene mouse melanoma (#937312)"
 /tissue_type="melanoma"
 /dev_stage="M2 cells"
 /lab_host="SOAR (kanamycin resistant)"
 /note="Organ: Skin; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. From M2 cells; a highly metastatic derivative of
 the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG
 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT 139 a 81 c 109 g 122 t 2 others
 ORIGIN

Query Match 87.0%; Score 20; DB 45; Length 453;
 Best Local Similarity 78.3%; Pred. No. 14;

Matches 18; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 OY 1 AARATGATGGATCTTCACAC 23
 Db 389 AARATGATGGATCTTCACAC 411

RESULT 13
 LOCUS AA298585 343 bp mRNA EST 18-APR-1997
 DEFINITION EST114216 Hsc172 cells II Homo sapiens cDNA 5' end similar to
 phosphatidylinositol 3-kinase, alpha catalytic subunit, mRNA
 sequence.

ACCESSION AA298585
 MID g1950938
 VERSION AA298585.1 GI:1950938
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 343)
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fullmer,R.A.,
 Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C.,
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fife,L.D.,
 Fitzgerald,L.M., Flygden,W.M., Fritchman,J.L., Geoghegan,N.S.,
 Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,P.T., Pellicciolo,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,G.A.,
 Dinko,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
 He,M.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.R.,
 Kozak,D.L., Kunsch,C., Hungjun,T., Li,H., Weissner,P.S., Olsen,H.,
 Raymond,L., Wel,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
 Dillion,P.J., Fannon,M.R., Kosen,C.A., Haseltine,W.A., Fields,C.,
 Fraser,C.M., and Venter,J.C.
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl), 3-174 (1995)

TITLE JOURNAL
 MEDLINE
 COMMENT

On Sep 12, 1996 this sequence version replaced gi:1392803.
 Other ESTs: THCA68479
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavage@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES
 source 1..343
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):1179595"
 /db_xref="taxon:9606"
 /clone_lib="HSC172 cells II"
 /cell_type="fibroblast"
 /dev_stage="fetal"
 /note="Organ: Lung; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI"

BASE COUNT 132 a 61 c 64 g 84 t 2 others
 ORIGIN

Query Match 82.6%; Score 19; DB 31; Length 343;
 Best Local Similarity 73.9%; Pred. No. 39;

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Cyt nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:13 ; Search time 539.84 Seconds
(without alignments)
206.192 Million cell updates/sec

Title: US-09-325-095-14

Perfect score: 35
Sequence: 1 GATGATGGCCARCTGTTTCATATGATGATTTGGCCA 35

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database:

GenBank: 1: gb_bal: 2: gb_ba2: 3: gb_om: 4: gb_ov: 5: gb_pat: 6: gb_ph: 7: gb_p11: 8: gb_p12: 9: gb_p13: 10: gb_p14: 11: gb_p15: 12: gb_p16: 13: gb_p17: 14: gb_p18: 15: gb_p19: 16: gb_p20: 17: gb_p21: 18: gb_p22: 19: gb_p23: 20: gb_p24: 21: gb_p25: 22: gb_p26: 23: gb_p27: 24: gb_p28: 25: gb_p29: 26: gb_p30: 27: gb_p31: 28: gb_p32: 29: gb_p33: 30: gb_p34: 31: gb_p35: 32: gb_p36: 33: gb_p37: 34: gb_p38: 35: gb_p39: 36: gb_p40: 37: gb_p41: 38: gb_p42: 39: gb_p43: 40: gb_p44: 41: gb_p45: 42: gb_p46:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	38	5	A37242 Sequence 11

2	31.4	89.7	3452	4	AF001076	AF001076 Gallus ga
3	31.4	89.7	3389	17	AF001075	AF001075 Avian sar
4	29.8	85.1	3207	3	BOVPHOS3KN	M93252 Bovine phop
5	29.8	85.1	3498	5	A37232	A37232 Sequence 1
6	28.2	80.6	3424	10	HSPH13K	U29090 H.sapiens m
7	28.2	80.6	3207	10	HSU79143	U03279 Human phosp
8	28.2	80.6	3207	12	MMU03279	U03279 Mus musculu
9	26.6	76.0	3868	10	HSP110DEL	Y10055 H.sapiens m
10	26.6	76.0	3525	10	HSU57843	U57843 Human phosp
11	26.6	76.0	5220	11	HSU86453	U86453 Human phosp
12	25	71.4	3132	12	MMU86587	U86587 Mus musculu
13	24.8	70.9	3213	10	SE7334	SE7334 phosphatidy
14	24.6	70.3	3712	36	DMPHOS3XT	Y09070 D.melanog
15	23.8	68.0	3213	12	RMD012482	AC002292 Rattus no
16	23	65.7	120787	8	CR097662	U97662 Chlamydomon
17	22.8	65.1	2200	10	HSC2P13K	Y11312 H.sapiens m
18	22.6	64.6	7654	10	HSC2P13K	Y13882 Homo sapien
19	22.6	63.4	5296	10	HSP13K1N	AC004199 Homo sapi
20	22.2	63.4	40622	11	AC004199	U53588 Homo sapien
21	22.2	63.4	37160	11	HSU53588	U34478 Dictyostell
22	22.2	63.4	4758	36	DDU23478	AC005543 Homo sapi
23	22.2	63.4	79810	42	AC005543	A17975 Glucose-6-p
24	21.8	62.3	1696	5	A17975	AC005896 Arabidops
25	21.8	62.3	119001	8	ATAC005896	AC005384 Homo sapi
26	21.8	62.3	178347	11	AC005384	AC006442 Homo sapi
27	21.4	61.1	155935	11	AC006442	AC003026 Human Chr
28	21.4	61.1	68355	11	HUAC03026	AC005636 Drosophil
29	21.4	61.1	70702	35	AC005636	AJ000008 Homo sapi
30	21.2	60.6	4880	9	HSC2P13K1	AB009636 Rattus no
31	21.2	60.6	5990	12	AB009636	AC007771 Drosophil
32	21.2	60.6	62980	35	AC007771	AC005555 Drosophil
33	21.2	60.6	101110	37	AC005555	D37799 Bacillus su
34	21.2	60.6	8054	1	BACAMOR000	AFO12285 Bacillus su
35	21	60.0	208230	1	BSUB00008	A47315 Sequence 6
36	21	60.0	46864	2	AF012285	X79944 Human Papil
37	21	60.0	662	5	A47315	X82892 D.melanog
38	21	60.0	612	17	HOV5731	DB3984 Sulciculus di
39	21	60.0	7215	36	DMP13K6BD	U52192 Drosophila
40	21	60.0	15329	36	SUCIDOLM	Z46973 H.sapiens A
41	21	60.0	6831	37	DMU52192	M73514 Influenza A
42	20.8	59.4	2990	10	HSP1R1	X74468 Human papil
43	20.8	59.4	2341	17	FLAHANP2	AC006397 Homo sapi
44	20.8	59.4	7412	17	HPV15	
45	20.8	59.4	91460	42	AC006397	

ALIGNMENTS

RESULT 1
A37242 38 bp DNA PAT 05-MAR-1997
LOCUS Sequence 11 from Patent WO9403609.
DEFINITION A37242
ACCESSION 92294353
VERSION A37242.1 GI:2294353
KEYWORDS
SOURCE
ORGANISM
unclassified.
unclassified.
unclassified.

REFERENCE 1 (bases 1 to 38)
AUTHORS Parker,P.T., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
TITLE EUCHARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 11 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
source Location/Qualifiers
1..38
/organism="unclassified"
/db_xref="taxon:32644"

BASE COUNT 7 a 6 c 8 g 11 t 6 others
ORIGIN

Query Match 94.3%; Score 33; DB 5; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 1 GATGATGGCCACTGTTTCATATGATTTGGCCA 35
|||||
1 1 GATGATGGCCACTGTTTCATATGATTTGGCCA 35

RESULT 2
AF001076 3452 bp mRNA VRT 08-JUL-1997
LOCUS
DEFINITION Gallus gallus phosphoinositide 3-kinase catalytic subunit mRNA,
complete cds.
ACCESSION AF001076
NID 92245505
VERSION AF001076.1 GI:2245505
KEYWORDS
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 3452)
AUTHORS Chang,H.W., Aoki,M., Fruman,D., Auger,K.R., Bellacosa,A.,
Tschlis,P.N., Cantley,L.C., Roberts,T.M. and Vogt,P.K.
TITLE Transformation of chicken cells by the gene encoding the catalytic
subunit of PI 3-kinase
JOURNAL Science 276 (5320), 1848-1850 (1997)
MEDLINE 97334438
REFERENCE 2 (bases 1 to 3452)
AUTHORS Chang,H.W., Aoki,M. and Vogt,P.K.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
BCC-239, La Jolla, CA 92037, USA
location/Qualifiers

FEATURES
source
1. 3452
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="brain"
1. 150
151..3354
/gene="c-p3k"
/note="p3k proto-oncogene"
151..3357
/note="c-p3k"
/codon_start=1
/product="phosphoinositide 3-kinase catalytic subunit"
/protein_id="AAB62534.1"
/db_xref="PDB:92245506"
/db_xref="GI:2245506"
/translation="MPPRSSGELGIMHPRIIVLECLLPMGIVTLLECLREATLLT
IKHELEKARKYPLVQLDDESSYIEFVSQEARREFETRLCDLRFQPLVI
EPNGNEEKILNREIGFAIGMPCIEFDVADPEQDPRRLIIVKCEAVDLRANPH
SRLLYCPVWESSPELPRIVKLDKGOIIVIVIVISNNDKQYTLKINDCPE
QYAAIRKTRSMILSSBOLKCVLEIQKILKVOGCEVLEKTPISQIVINS
IMGRPNMLMAKESLYTQPLDTTPMYSRISTATYNGEATASLWINSAL
RRLICATYVNNIRIDIKYVGTGTYHGEPLCDVNNQTRVCSNPRNEMTSYMY
IPDLPRARCLSTICSVKGRGAKEECPPLAMGNINMFDYDTLVGKMAINLMAVH
GLEDLNPITGVGNSNPKEPCLEFEDMFSNPKFPDMVIEEHNMTISRELGNV
SYAGLSNRILARDELRESDEQRLAICTRPLSEITQERDIFMSHRKICVNPPELP
KLLSVKWSRDEVAQMTCLVNDPPIKEQAMELDCNPDVAVFAVRCLEKLT
DDLSQYLQLVQVLYEQLDQVLEFLKALITQRIQGHFFMHLKSMHKKTVSQ
RFGLEISYCRACGMVYKHLISROYEAMEKILNLDILKQEKDETQKVMKFLVEQMR
RPFMDALQGIISPLNPAHOLGNLRECRINSKAPPLAMENPDIEMSELFFONNE
IEFKNGDDLRODMTLQIIRIMENIMONOGIDRLMNYGCLSTGDCVGLLEVRSSTH
INQIQCKGGLKALQFNSTHLMWLDKNGEYDAIDLFTSCAGYCATITLISIG
DRNNSINVKDCQGLHIDGHFLDKHKKKFGYKRRKRVPLTQDLVYISKADQCT
KREFREFQEMCKRAYLAIQHNFLFINFSMLGSGMPELQSFDDIAVIRKTLADK
TEQEALEYFMKQNDNAHGGWTKMDMIETIKOHALN"
3358..3452

3'UTR

BASE COUNT 1073 a 672 c 783 g 924 t
ORIGIN

Query Match 89.7%; Score 31.4; DB 4; Length 3452;
Best Local Similarity 82.9%; Pred. No. 0.003;
Matches 29; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

1 1 GATGATGGCCACTGTTTCATATGATTTGGCCA 35
|||||
1 1 GATGATGGCCACTGTTTCATATGATTTGGCCA 2957

Db 2923 GATGATGGCCACTGTTTCATATGATTTGGCCA 2957

RESULT 3
AF001075 3389 bp mRNA VRL 08-JUL-1997
LOCUS
DEFINITION Avian sarcoma virus 16 gag-v-phosphoinositide 3-kinase catalytic
subunit fusion protein (gag-v-p3k) and retroviral env mRNAs,
partial cds.
ACCESSION AF001075
NID 92245502
VERSION AF001075.1 GI:2245502
KEYWORDS
SOURCE Avian sarcoma virus 16.
ORGANISM Avian sarcoma virus 16.
Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
REFERENCE 1 (bases 1 to 3389)
AUTHORS Chang,H.W., Aoki,M., Fruman,D., Auger,K.R., Bellacosa,A.,
Tschlis,P.N., Cantley,L.C., Roberts,T.M. and Vogt,P.K.
TITLE Transformation of chicken cells by the gene encoding the catalytic
subunit of PI 3-kinase
JOURNAL Science 276 (5320), 1848-1850 (1997)
MEDLINE 97334438
REFERENCE 2 (bases 1 to 3389)
AUTHORS Chang,H.W., Aoki,M. and Vogt,P.K.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
BCC-239, La Jolla, CA 92037, USA
location/Qualifiers

FEATURES
source
1. 3389
/organism="Avian sarcoma virus 16"
/db_xref="taxon:60629"
1. 3252
/gene="gag-v-p3k"
/note="fusion gene of gag and v-p3k"
1..78
/gene="gag-v-p3k"
/note="derived from gag gene"
1..3252
/gene="gag-v-p3k"
/codon_start=1
/product="gag-v-phosphoinositide 3-kinase catalytic
subunit fusion protein"
/protein_id="AAB62532.1"
/db_xref="PDB:92245503"
/db_xref="GI:2245503"
/translation="QGLIRTPSLITTPGEIIRKYVDROKASHLMPRIIVLECLLPG
MVTLECEAEALTLIKHELEKARKYPLVQLDDESSYIEFVSQEARREFETRL
RCDLRLFPFLKVIIEPVNGNEEKILNREIGFAIGMPCIEFDVADPEQDPRRLIN
VKEAVDLRANPHSRLLYCPVWESSPELPRIVKLDKGOIIVIVIVISNNDKQYTLKINDCPE
QYAAIRKTRSMILSSBOLKCVLEIQKILKVOGCEVLEKTPISQIVINS
IMGRPNMLMAKESLYTQPLDTTPMYSRISTATYNGEATASLWINSAL
RRLICATYVNNIRIDIKYVGTGTYHGEPLCDVNNQTRVCSNPRNEMTSYMY
IPDLPRARCLSTICSVKGRGAKEECPPLAMGNINMFDYDTLVGKMAINLMAVH
GLEDLNPITGVGNSNPKEPCLEFEDMFSNPKFPDMVIEEHNMTISRELGNV
SYAGLSNRILARDELRESDEQRLAICTRPLSEITQERDIFMSHRKICVNPPELP
KLLSVKWSRDEVAQMTCLVNDPPIKEQAMELDCNPDVAVFAVRCLEKLT
DDLSQYLQLVQVLYEQLDQVLEFLKALITQRIQGHFFMHLKSMHKKTVSQ
RFGLEISYCRACGMVYKHLISROYEAMEKILNLDILKQEKDETQKVMKFLVEQMR
RPFMDALQGIISPLNPAHOLGNLRECRINSKAPPLAMENPDIEMSELFFONNE
IEFKNGDDLRODMTLQIIRIMENIMONOGIDRLMNYGCLSTGDCVGLLEVRSSTH
INQIQCKGGLKALQFNSTHLMWLDKNGEYDAIDLFTSCAGYCATITLISIG
DRNNSINVKDCQGLHIDGHFLDKHKKKFGYKRRKRVPLTQDLVYISKADQCT
KREFREFQEMCKRAYLAIQHNFLFINFSMLGSGMPELQSFDDIAVIRKTLADK
TEQEALEYFMKQNDNAHGGWTKMDMIETIKOHALN"
3358..3452

SCAGCVATFLIGDRHNSINWKDDGOLFHDHGHLDHKKKKFGYKRRRPPVLT
ODLIIVSKAOBECTKREPERROEMCHKAYLAROMANPEFINFSGMGLSPELOS
FDDIATIRKTLADKITEOEALEIFMKOMNDAHGCGWTKMDWIFHTIKOHALN"
79. .84
/gene="gag-v-p3k"
/note="sequence of unknown origin"
85. .3252
/gene="gag-v-p3k"
/note="derived from c-p3k gene"
3250. .3370
3371. .>3389
/codon_start=1
/product="retroviral env"
/protein_id="AA82533.1"
/db_xref="PID:g2245504"
/db_xref="GI:2245504"
/translation="MARNS"
BASE COUNT 1071 a 655 c 751 g 912 t
ORIGIN

Query Match 89.7%; Score 31.4; DB 17; Length 3389;
Best Local Similarity 82.9%; Pred. No. 0.003;
Matches 29; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Cy 1 GATGATGCCACRGTTCATATGATTTGGCCA 35
||||| ||:||||:||||:|||||
Db 2818 GATGATGACACACTTCATATGATTTGGCCA 2852

RESULT 4
LOCUS BOVPHOS3N 3207 bp mRNA MAN 19-AUG-1992
DEFINITION Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete cds.
ACCESSION M93252
NID 9163519
VERSION M93252.1 GI:163519
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE Bos taurus cDNA to mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovines; Bos.
1 (bases 1 to 3207)
Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R.,
Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Jotty, N.F.,
Hsuan, J.J., Courtneidge, S.A., Parker, P.J. and Waterfield, M.D.
Phosphatidylinositol 3-kinase: Structure and expression of the 110
kd catalytic subunit
Cell 70, 419-429 (1992)
92354059 Location/Qualifiers
source
1. .3207
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBAR-1"
1. .3207
/EC_number="2.7.1.67"
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="AA30698.1"
/db_xref="PID:g163520"
/db_xref="GI:163520"

CDS
1..3207
/translation="MPPRPSGGLMGILHMPRLIVGCLLPGNMIIVTLDECUREATLIT
IKHELFEARKYPLHOLLDESSYIFVSVQEARREFEDETRRLCDRLRQPLKAYI
EPVGNREKILNREIGFALIGMPVEFPMVNDPEVDNRRLNLYKRAYDRLDINSPE
SRNAVYVPNVESPELPKRLYNLDGQILVIVTWYVSPNDKQKTLTKNHCVPE
QVIAEAIKRTKRLSLSEQLCLVLEQYILKVGCDYFLKPLSOKYIRSC
IMGRMNLIMAKESLYSOLPMDCFMPSYSRISATPYMNGETSKSLMYNSAL
RIKLICATYVAVNRIDIKIYRTGIGHGEPICDNNVTOFPCSNPRMNLNLYDI
IPDLPRARLICSTICSVKGRKAKEHCPKAMGNINFDYDTLVSGMALNLPVPH
GLEDLNPIGYTGSNPKETPCLELEFDWSSVYKFPDMASIEEHANWSVRENGFSY

SHAGLSRLAARDNLRENDKREQLAICTRDLISITRDEKDFILSHRHVCYTIPEILP
KLLSVKWNNSDEVAQMTCLVKDMPPIKPEQAMBLDLCNPPDPVRFPAVACLKYL
DDKLSOYLQIOLVQYLYEQLYDNLVLEFKLTKALTNORIGFEFHLKSEHNNTVSQ
RFGLLSYSCACGMYLKHLNRQVEAEKLNLDIDIOEKDEKIQVOMKFLVEQNS
RPDMDALOGFLSPINPAHQGNRLRECRIMSSAKRPLNMENPDIMSELLRONS
IIFKNGDILRODMTLTQIRIMENIMONOGIDMLPYGCLSTGDCGLIEVYNSHT
IMQIOCKGKGLGALQFNSHTLHQLKRNKGEIYDAIDIFTRSCAGCVATFLIGIG
DRHNSINWKDDGOLFHDHGHLDHKKKRGYKRRRPPVLTODFLIVSKAOBECT
KTRFEPEQEMCYKAYLAROMANPEFINFSGMGLSPELOSFDIATIRKTLADK
TEOEALEIFMKOMNDAHGCGWTKMDWIFHTIKOHALN"
BASE COUNT 1028 a 581 c 680 g 918 t
ORIGIN

Query Match 85.1%; Score 29.8; DB 3; Length 3207;
Best Local Similarity 80.0%; Pred. No. 0.014;
Matches 28; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Cy 1 GATGATGCCACRGTTCATATGATTTGGCCA 35
||||| ||:||||:||||:|||||
Db 2773 GATGATGACACACTTCATATGATTTGGCCA 2807

RESULT 5
LOCUS A37232 3498 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9403609.
ACCESSION A37232
NID 92294345
VERSION A37232.1 GI:2294345
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3498)
AUTHORS Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
TITLE EUCARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL IMP CANCER RES TECH (GB)
COMMENT Other publication Jp 8503124T 960409.
FEATURES
source
1. .3498
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
1. .3207
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PID:e304808"
/db_xref="GI:2294346"
/db_xref="PID:g2294346"
/db_xref="GI:2294346"
/translation="MPPRPSGGLMGILHMPRLIVGCLLPGNMIIVTLDECUREATLIT
IKHELFEARKYPLHOLLDESSYIFVSVQEARREFEDETRRLCDRLRQPLKAYI
EPVGNREKILNREIGFALIGMPVEFPMVNDPEVDNRRLNLYKRAYDRLDINSPE
SRNAVYVPNVESPELPKRLYNLDGQILVIVTWYVSPNDKQKTLTKNHCVPE
QVIAEAIKRTKRLSLSEQLCLVLEQYILKVGCDYFLKPLSOKYIRSC
IMGRMNLIMAKESLYSOLPMDCFMPSYSRISATPYMNGETSKSLMYNSAL
RIKLICATYVAVNRIDIKIYRTGIGHGEPICDNNVTOFPCSNPRMNLNLYDI
IPDLPRARLICSTICSVKGRKAKEHCPKAMGNINFDYDTLVSGMALNLPVPH
GLEDLNPIGYTGSNPKETPCLELEFDWSSVYKFPDMASIEEHANWSVRENGFSY
SHAGLSNRLARDNLRENDKREQLAICTRDLISITRDEKDFILSHRHVCYTIPEILP
KLLSVKWNNSDEVAQMTCLVKDMPPIKPEQAMBLDLCNPPDPVRFPAVACLKYL
DDKLSOYLQIOLVQYLYEQLYDNLVLEFKLTKALTNORIGFEFHLKSEHNNTVSQ
RFGLLSYSCACGMYLKHLNRQVEAEKLNLDIDIOEKDEKIQVOMKFLVEQNS
RPDMDALOGFLSPINPAHQGNRLRECRIMSSAKRPLNMENPDIMSELLRONS
IIFKNGDILRODMTLTQIRIMENIMONOGIDMLPYGCLSTGDCGLIEVYNSHT
IMQIOCKGKGLGALQFNSHTLHQLKRNKGEIYDAIDIFTRSCAGCVATFLIGIG
DRHNSINWKDDGOLFHDHGHLDHKKKRGYKRRRPPVLTODFLIVSKAOBECT
KTRFEPEQEMCYKAYLAROMANPEFINFSGMGLSPELOSFDIATIRKTLADK
TEOEALEIFMKOMNDAHGCGWTKMDWIFHTIKOHALN"


```

RESULT 10
LOCUS HS057843
DEFINITION Human phosphatidylinositol 3-kinase delta catalytic subunit mRNA, complete cds.
ACCESSION U57843
VERSION 92076750
KEYWORDS 057843.1 GI:2076750
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 3525)
AUTHORS Mahlum,C.E., Becker,R.P. and Morris,A.J.
TITLE H. sapiens mRNA for phosphoinositide 3-kinase delta catalytic subunit
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3525)
AUTHORS Morris,A.J.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1996) Andrew J. Morris, Pharmacology, SUNY-Stony Brook, BSR, Level 7 Room 168, SUNY-HSC, Stony Brook, NY 11794-8651, USA

FEATURES
Source Location/Qualifiers
1..3525
/organism="Homo sapiens"
/db_xref="taxon:9606"
116..3322
/codon_start=1
/product="phosphatidylinositol 3-kinase delta catalytic subunit"
/protein_id="AA053966.1"
/db_xref="PI-3 kinase catalytic subunit"
/translacion="MPGVDCPMEFMTKEENOSVYVDELLPTGYVLPVSRNANLST
IKOLLHNAOYEPFLFHMLSAPKAYFTCTNOTAEQOELDEDRRLCDVOPPIYARLY
AREGDRVKKLINSQSLILGKGLHEBDSLCDEPVNDRFRKMGCFCEBAARQQLGME
AWLOYSFPLQLEPSAQTWGPGLRLPNRALLVNAKFESESEFTFOVSTKQVAPALMA
CALRKATVFRQPLVEQPEDYTLQVNGHELYGNYPICQPOVORPRAPPIIPARK
SSVLSMSLEQPERIELIOGSKYNADEMRKLVVQAGLFGNEMLCKTVSSSEVSKSP
VKORLEPFINICDLPBMAKLCFALYAVTEKAKKARSTKSKADCPYAMNMLLFD
YDOLKIGERCILYMPSPVDEGELLNPTGYRNSPNDASALILCLPEVAPPHYVY
PALEKUPGMLASREMRPELMKRLHEDSPSRNIGVAMGGLVLELVAPGDAB
CSRLQILELGRHSECVHTEEQQLREILERRSGELYEHEKDLVWKLHREVEHEF
PEALARKLLVTVNMNHEVAPDALPAVSWPELVSLELIDFSEPDCHVGSFAIKSL
RKLTDDELFOYLQVLYQVLYKESYLDCETKFLERALLNRKIGLFWHLSEHMP
SVLRPEGLLEAYRGSTHMMVLMKOGELSKLALNDPVLSSOKTPQPKTELMH
LCMRQEVYEAISHLOSPLDSTLAEVQECSTPMDSKMRPIAMTNSNEEGSGGTV
GIIFKNGDYLDRDMTLQITQIMDYLKOBGLDRLKRTPIPGCLPTGDRGLIYVLRSD
TTANIOUNKSNAATAFAENKDALVNMKSKNGEALDRATIEFTLSCAGYCAVATVIG
IGDRSDNIMRESQGLFIDGHLGNFKTFGINREVRPILYDVLHVLYQOKTN
NSEKFERGCEYATYTLRRHGLFLHLFALMRAAGLPELSCSKDIOYLKNDSLAGK
TEEALKHREKFNALREGWTKYVMNLAHNYSKDNRO"
BASE COUNT 761 a 1056 c 1067 g 641 t
ORIGIN

Query Match 76.0%; Score 26.6; DB 10; Length 3525;
Best Local Similarity 74.3%; Pred. No. 0.32;
Matches 26; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

```

ACCESSION U86453
MIM 92317893
VERSION U86453.1 GI:2317893
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 5220)
AUTHORS Chantry,D., Vojtek,A., Kashishian,A., Holtzman,D.A., Wood,C.,
Gray,P.W., Cooper,J.A. and Hoekstra,M.F.
TITLE p10delta, a novel phosphatidylinositol 3-kinase catalytic subunit
that associates with p85 and is expressed predominantly in
leukocytes
JOURNAL J. Biol. Chem. 272 (31), 19236-19241 (1997)
MEDLINE 97382246
REFERENCE 2 (bases 1 to 5220)
AUTHORS Chantry,D., Vojtek,A., Kashishian,A., Holtzman,D., Wood,C.,
Gray,P.W., Cooper,J.A. and Hoekstra,M.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Leukocyte Biochemistry, ICOS Corporation,
22021 20th Ave. S.E., Bothell, WA 981021, USA

FEATURES
Source Location/Qualifiers
1..5220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="macrophage"
196..3330
/note="PI-3 kinase catalytic subunit"
/codon_start=1
/product="phosphatidylinositol 3-kinase catalytic subunit p10delta"
/protein_id="AA025677.1"
/db_xref="PI-3 kinase catalytic subunit"
/translacion="MPGVDCPMEFMTKEENOSVYVDELLPTGYVLPVSRNANLST
IKOLLHNAOYEPFLFHMLSAPKAYFTCTNOTAEQOELDEDRRLCDVOPPIYARLY
AREGDRVKKLINSQSLILGKGLHEBDSLCDEPVNDRFRKMGCFCEBAARQQLGME
AWLOYSFPLQLEPSAQTWGPGLRLPNRALLVNAKFESESEFTFOVSTKQVAPALMA
CALRKATVFRQPLVEQPEDYTLQVNGHELYGNYPICQPOVORPRAPPIIPARK
SSVLSMSLEQPERIELIOGSKYNADEMRKLVVQAGLFGNEMLCKTVSSSEVSKSP
VKORLEPFINICDLPBMAKLCFALYAVTEKAKKARSTKSKADCPYAMNMLLFD
YDOLKIGERCILYMPSPVDEGELLNPTGYRNSPNDASALILCLPEVAPPHYVY
PALEKUPGMLASREMRPELMKRLHEDSPSRNIGVAMGGLVLELVAPGDAB
CSRLQILELGRHSECVHTEEQQLREILERRSGELYEHEKDLVWKLHREVEHEF
PEALARKLLVTVNMNHEVAPDALPAVSWPELVSLELIDFSEPDCHVGSFAIKSL
RKLTDDELFOYLQVLYQVLYKESYLDCETKFLERALLNRKIGLFWHLSEHMP
SVLRPEGLLEAYRGSTHMMVLMKOGELSKLALNDPVLSSOKTPQPKTELMH
LCMRQEVYEAISHLOSPLDSTLAEVQECSTPMDSKMRPIAMTNSNEEGSGGTV
GIIFKNGDYLDRDMTLQITQIMDYLKOBGLDRLKRTPIPGCLPTGDRGLIYVLRSD
TTANIOUNKSNAATAFAENKDALVNMKSKNGEALDRATIEFTLSCAGYCAVATVIG
IGDRSDNIMRESQGLFIDGHLGNFKTFGINREVRPILYDVLHVLYQOKTN
NSEKFERGCEYATYTLRRHGLFLHLFALMRAAGLPELSCSKDIOYLKNDSLAGK
TEEALKHREKFNALREGWTKYVMNLAHNYSKDNRO"
BASE COUNT 1120 a 1525 c 1483 g 1092 t
ORIGIN

Query Match 76.0%; Score 26.6; DB 11; Length 5220;
Best Local Similarity 74.3%; Pred. No. 0.34;
Matches 26; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

```

RESULT 12
LOCUS MM086587
DEFINITION Mus musculus phosphatidylinositol 3-kinase catalytic subunit p110
ACCESSION U86587

```

NID	VERSION	Q2331237	Q2331237
KEYWORDS		086587.1	GI:2331237
SOURCE		house mouse.	
ORGANISM		Mus musculus.	
REFERENCE		Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS		1 (bases 1 to 3132)	
TITLE		Chanly, P.W., Voitek, A., Kashishian, A., Holtzman, D.A., Wood, C., Gray, P.W., Cooper, J.A., and Heckstra, M.F.	
JOURNAL		plilodelta, a novel phosphatidylinositol 3-kinase catalytic subunit that associates with p85 and is expressed predominantly in leukocytes	
MEDLINE		J. Biol. Chem. 272 (31), 19236-19241 (1997)	
REFERENCE		2 (bases 1 to 3132)	
AUTHORS		Voitek, A.B. and Cooper, J.	
JOURNAL		Direct Submission	
FEATURES		Submitted (22-JAN-1997) Biological Chemistry, University of Michigan, 3323 MSRB III, Box 0636, Ann Arbor, MI 48109-0636, USA	
SOURCE		Location/Qualifiers	
CDS		1..3132	
		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		/tissue_type="spleen"	
		1..3132	
		/codon_start=1	
		/product="phosphatidylinositol 3-kinase catalytic subunit p110 delta"	
		/protein_id="AAC55676.1"	
		/db_xref="PID:g2331238"	
		/db_xref="GI:2331238"	
		/translation="MPPEVDDPMEFWTKESQSVVDFLLPTGVYLFPPVSRNLTSTIKYLAHRAQYEPFLPMISDEPAVYFCVNTAQOELDEDRICDQPLPYRLALGLWYAREDDRYKLINSOISLIAGLHEPDSLRPEVNDFTKRCOEBAARLOGLWYEMOYSPLEPESAPGRAGLLRVSNRALLVYKFPSESESTPQVSTKMPALANWCALEKRAKTVROPDYVEQPEERIALDYNGHETLYGNYPCFCHFOYICSCISHSLETHLTVHSSILAMRDEOSNPAPQVQKPRAPKPPYIPAKRPSVSLMSDPESTILEGRKYNVADESMKLVQAAGLEHNGEMLCRTVSSSEVNVCSPPWYKQLEFPISYVDLPRMRLCEALYVVEKAKRKARSTKRSKADCPILMANMLFDYKDLQTLGRCILAMPSVDEKGGELINAPAGVGRNPNTESAAALVITLPEVAPRYVEPALKEITLGRGERGRITEEELOLRPIELRRSGELYEHRKDLVYKMHKRVHOPPEALARLLLYTKMKHEDVADMYILCSPELPEVLSALTELLBSFPDCYVGSFALSLKRLDDELFOYLLQVYKTESYLDCESTKFLGALANRRIGHLEFWHLSESNVPSVALRPGILMEAYCGSHHAKVYLMKOGALSRLKALNDNFVYSQKTKRTPQTKEMHNCMRQETMEALSHLOSIPDSTLIDLEEVEQCEQTEFMSCKRPLMYISEEGSAGNVGILFENGDDLQDMUTLQIMDLVLMKREGDLNMTPYGCLPTGDRGTLIEVILSPDILANTOLKSNMAATAANKRALIKWAKNKGELADRAIEBFTLSCACVATYVLYGGRDSNDINRESQGLHPIRGEHNLGNKTKFGLINRERVPFTLYTDFVHVAVIQQGTNNSEKFERRGVCERYVTLIRRHGDLFLHLFALMRAGLDELSCSDIOYLKIDSLAGKTEEBALKHFRVKAFFNALRESWKTKVYNWLAHNSKDRQ"	
BASE COUNT		690 a 920 c 901 g 621 t	
ORIGIN			
Query Match		71.4%; Score 25; DB 12; Length 3132;	
Best Local Similarity		71.4%; Pred. No. 1.5;	
Matches		25; Conservative 5; Mismatches 5; Indels 0; Gaps 0;	
Db		2704 GAGACTGGCAGACTTCACATGATTTGGCCA 2738	
RESULT		1 CATGATGCCARCTGTTTCATATGCATTTGGCCA 35	
LOCUS		S67334	
DEFINITION		phosphatidylinositol 3-kinase p110 beta isoform-110 kda catalytic subunit [human, mRNA Partial, 3213 nt].	
ACCESSION		S67334	
NID		9455759	
VERSION		S67334.1	
KEYWORDS		GI:455759	

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS 1 (bases 1 to 3213)
TITLE Hu.P., Mondino, A., Skolnik, E.Y. and Schlessinger, J.
Cloning of a novel, ubiquitously expressed human
phosphatidylinositol 3-kinase and identification of its binding
site on p85
Mol. Cell. Biol. 13 (12), 7677-7688 (1993)

JOURNAL
MEDLINE 94067128
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gidsbg_140879] from the original journal article.
This sequence comes from Fig. 1.

FEATURES
source location/Qualifiers
1..3213
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3213
/gene="phosphatidylinositol 3-kinase p110 beta isoform, PI
3-kinase p110 beta"
1..3213
/gene="phosphatidylinositol 3-kinase p110 beta isoform, PI
3-kinase p110 beta"
/note="110 kda catalytic subunit; This sequence comes from
Fig. 1; PI 3-kinase p110 beta"
/codon_start=1
/product="phosphatidylinositol 3-kinase p110 beta isoform"
/protein_id="AA829081.1"
/db_xref="PID:g455760"
/db_xref="GI:453760"
/translation="MCFSITMPAMADIDIMAVDSQIASGDISPVDFLLPTGIYIQGL
EVRERATISYIKOMLKVQVNPENFLNMDIDSMEACVMQIAYELEDETRLCQV
RPLPVLKIVLRSGDKSKIGVIGLGHPELDLPDENVEFRKKRKFESEKE
LSLVGSWMDMIKOTYPRHEPSIPHEPSIPHEPSIPHEPSIPHEPSIPHEPSIP
NPKFVNELAQKRILTIKGEDVSPHYTVQVSGRYVYVGDPILIOFYIRNCYNMR
ALPFILVECCKIIRKTEQEMIAIEAINSSMLPLPLPPKRTRIISHWENNPFQ
IVLVGRKLTEETVRAKRVAGLEHGELLCTIVSSVGRNDHIWELEDINIC
DEPMARCLFAVAVLDVKVTKSTKINSKQTIIRAKGVHPVAVMTWVEDFGNQ
QIRGTIIILHSMSFPDELMEANLMPTQTOINPTENATLHVAFENKROOPYYPPE
DLTERKAETILASSDSANVSVSRGGKKPLPVKTELIDRPLOSLOENENDLIWTLRQCR
ETIPQSIPRLILSIKKMKLEDVAQLAIDLINPRLPREALELIDFNVPQYYREYAV
GICRQNSDELSQILLQVOYLATEPLDCALSFFLERALGNRRIGQFLPHURSEV
HIPVSVQFVITLAEYRGVSVGHMKVISKOVELANKTILNSLIKLAVALNRAGEE
AMVTICLKOSAYREALSDIOSLPNCPVILSELLEYEKCYKDSKKPLVLYNNKFGE
SVAGTICRGGDDLRODMTLTMLRMJLMDLMEAGLDMLPYGCATGDRGLEVEST
SETADIOLSNSSNVAAAAPNRKALLMKLEKYSGDIDSLIEFTTSICGYCVAAYV
LGIDTRSDININAKTIQOLFHDIFGHILNGFKSFEGIKREKRVPIILTYDTIHYQQSK
TGNETKRGFRQCCEDAYLLIRRGNTITFLPALMTAGLEPILTVDIYLDKQIAL
GKSSEBALKOFKFDLEALRESYTCKNMMAHYTRKYDRS"

BASE COUNT 979 a 612 c 704 g 918 t

ORIGIN

Query Match 70.9%; Score 24.8; DB 10; Length 3213;
Best Local Similarity 76.7%; Pred. No. 1.9;
Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 6 TGGCGACGTCTTCACATATGCAATTTCGCCA 35
|||||:|||||:|||||:|||||:|||||:
Db 2790 TGGCGACGTCTTCACATGACTTTGGACA 2819

RESULT 14
DMPHOS3KI
LOCUS DMPHOS3KI 3712 bp mRNA INV 07-JAN-1997
DEFINITION D.melanogaster mRNA for phosphoinositide 3-kinase, p110.
ACCESSION Y09070
VERSION 1
KEYWORDS Y09070.1 GI:1707447
phosphoinositide 3-kinase.
fruit fly.
SOURCE Drosophila melanogaster

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS

Eukaryote: Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; 1 (bases 1 to 3712)
Leewers, S.J., Weinkove, D., MacDougall, L.K., Hafen, E. and Materield, M.D.
The Drosophila phosphoinositide 3-kinase Dp110 promotes cell growth
EMBO J. 15 (23), 6584-6594 (1996)
2 (bases 1 to 3712)
Leewers, S.J.
Direct Submission
Submitted (25-OCT-1996) S.J. Leewers, Ludwig Institute for Cancer Research, University College/Middlesex Hospital Branch, 91 Riding House Street, London, W1P 8BT, UK
Location/Qualifiers
1. .3712
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="III"
/dev_stage="eye /antennal imaginal disc"
/map="92E12-13 mu"
294. .3560
/note="protein also known as PI3K-92E"
/codon_start=1
/product="phosphoinositide 3-kinase, Dp110"
/protein_id="CAA0291.1"
/db_xref="PID:e283788"
/db_xref="PID:g1707448"
/db_xref="GI:1707448"
/db_xref="FlyBase:FBgn0015279"
/db_xref="SPTREML:P91634"
/translation="MNMNDRALAYVHQPKEETPEBEAPPCEKRESVNLKMKEMLNM VDIICLNGFLELRVPANTIOVIVENVNDKAKPGLVYVKEACEYGVGISFEN IEBYDITKRLSEVQPFGLISGERDITSESDYELTKVMKMGITIDRHRTGHS PEIDDFLYMTQCDNIELRSAYTMQORLLYEPLRLANSTKPELIRRHPTREPL IYVKNENDOSTFLSVNEODTPSLSTLOKMRSQMKMNDSTSDYLKSGRDEYL LGDYPLQPIXTQEMLSDSAPNVVLOSVALRSYIHNHNEQAMVTRP.PPKRTVAL HKSTSLMDKNTFOLTHSISVNFQKTRALKVGVAVCLYHGDRLCAQRSTDSPG NFDITFLNDLMQFDIOMRLPMTRICLIVFEVTKMSRSKSSNNNDLAKQVPPK NPLAWNTTTFDKDILRTGRHTLYTVADDIQSVFHPGLIIEPNPKEECALD LTFESGTVARYSEEVLOYADREOVNRLOLAGPEPKIKELKELMANTGADK IYVMDODRANRMRFRALDIRELPELSTIIHCYVWKRNDVADMTLAKOMPLISTE RSELDIYAPDPAYRRFARCLHFLDEDLILYLQVQAIHNESTLESDIYVFLLE RALNRQIRGHTFFMLRSEMQTSMTREGLLEVLTKGCKHHAAPLRKOLHVEKIK QGSLIAKSGKEVKTMLQDFLRDQNSAVFONIONPLNSFKSGVTPRCRVMSK MRPLWVFEVADVNASDVHIFKNGDILRQDMTLQVLRVMDLKRDKGDFPMNTYN CISEMSLAGEIVVRAETIANLOKEKGSATSPFKGSLISMLKHNKPADKLNA INFETLSGAGCYCATVYLVGADRSHDNIMKRNQOLFHDGHIIGHFKERKGLVRRR VPRVITHDYVYVINKGNDRSEKKEFCFQELCERAPLYLKRHCLILSLSWMISTGL PELSEKDLIYLRKELVLDYTEKARHFAKSEALANSWKTSLNASHNESKNNKO "

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS

beta isoform; catalytic subunit; phosphatidylinositol 3-kinase. Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 3213)
Mulder, H.
Direct Submission
Submitted (03-NOV-1998) Mulder H., Department of Cell and Molecular Biology at Lund University, Section for Molecular Signalling, Box 94, SE-221 00 Lund, SWEDEN
2 (bases 1 to 3213)
Mulder, H., Stenson Holst, L. and Degerman, E.
Phosphatidylinositol 3-kinase and activation of phosphodiesterase 3b in adipocytes
Unpublished
Location/Qualifiers
1. .3213
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/cell_type="primary adipocyte"
1. .3213
/codon_start=1
/product="phosphatidylinositol 3-kinase, catalytic subunit, beta isoform"
/protein_id="CAA10046.1"
/db_xref="PID:e1340135"
/db_xref="PID:g3850840"
/db_xref="GI:3850840"
/translation="NCFRSIMPRAMDTIDIVAVDSQASDSISVDPLPRTIQL EYPRKATISTIKOMLKQVIMPFMLMDIOTSYFACVQNAVTELEDERRLDLY RPLPVALKLVTSKCDPAERKLDKIGVLGKGLHEFDALDPEVNERRMKRFSEDKI OSVILSWMDWKHTYPPHEBSYLENLEKGLVYVAVHENSQDYFSQVSNL NPRIKHELAIQKRLATRGKEEBSPCDYLYSGRYEYVAGHPILTOFYIRNCVNR TLPHFLVACCKKKMYEOMETAIEAIRNSSSLPLP.PPKRTVTSIHWGNNNPFQ IYVKNKNTERTYVYVRAFLFHGTELLCTVYSESLSGNDHNMNQLPFDINIC DLPRARLCPAYAVIADKRYTKRTITPSTQITRKAGKHTPAAVNTVAFDEKIC QNSGVILHSWSEFDELEEMNPQVOTNPVEMATAIHKPEPNKOPYYPPE DKIEAAELIASDSANVSSRGKRLAVLKEILDRDPLSOLCEMNDLIMLRDPCR ENPOSLPKLLSISKNKLEDAVQALQIIMPKEPPREALELDEPNYDDQVREYAV GCLRQSDPELISYLLQVLYQYEFLLCALSRFLLEALNRRIGOLFPHMLRSEV HPVAVSIOGVILEACRSGSGHMYLSKOVANLKTNTLSLITNMLKLRAGKE AMHTCLKASRYREALSDOSPLNPLYTSELSEVERCRVDSKRPMLVYSNRAGEK AVGVIRKNSDDLRQDLTLQMLRLMDLMLKEAGLDRLMPSGLAGDYSGLLEVST SETIADIQNSNVAVATPAFNKDALNMLLKEKNSGDDLRALIEFTLSGACVAVSY LGIDASHNSINMYKTKGOLFHDGHIHLENFSKFGIKREVPFILTGYHVIQOGK TGNTERGGRPPQCCEDAPVILRRHGLFTFLPALMTAGLPELTSVKDIQYLDKSLAL GNSFEELAKQFPQKDEALRESMTQVNMMAHTVRKQVRS"

Search completed: September 13, 1999, 15:55:15
Job time: 4543 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:29 ; Search time 255.05 Seconds

(without alignments)
34.333 Million cell updates/sec

Title: US-09-325-095-14

Perfect score: 35

Sequence: 1 GATGATGCCARCTGTCATATGCAATTTGGCCA 35

Scoring table: IDENTITY_MUC

Searched: 31585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	38	1	Q59013
2	33	94.3	38	1	Q57019
3	29.8	85.1	3207	1	Q51155
4	29.8	85.1	3498	1	Q57012
5	28.2	80.6	3412	1	Q51156
6	26.6	76.0	3387	1	V15533
7	26.6	76.0	5220	1	V13340
8	26.6	76.0	5220	1	X15932
9	21.8	62.3	1696	1	Q20959
10	21.8	62.3	1455	1	T30577
11	21.4	61.1	3252	1	X24400
12	21	60.0	381	1	O57522
13	21	60.0	662	1	T03504
14	21	60.0	3202	1	T47520
15	21	60.0	6831	1	T80200
16	21	60.0	3316	1	V44859
17	20.8	59.4	2121	1	V11860
18	20.6	58.9	1735	1	T37149
19	20.2	57.7	1461	1	T30576
20	19.8	56.6	6693	1	V06570
21	19.6	56.0	934	1	T05843
22	19.6	56.0	5285	1	T80199
23	19.6	56.0	5061	1	V42920
24	19.6	56.0	4956	1	X13228
25	19.6	56.0	1500	1	X20016
26	19.6	56.0	1399	1	X20017
27	19.2	54.9	674	1	T03503
28	19.2	54.9	665	1	T03505
29	19.2	54.9	3602	1	V01877
30	19.2	54.9	2487	1	V01878
31	19.2	54.9	3324	1	V01879
32	19.2	54.9	2451	1	V01876
33	19.2	54.9	410	1	V15536
34	19.2	54.9	12565	1	V62352
35	19.2	54.3	674	1	T03506
36	19	54.3	1193	1	T79975
37	19	54.3	1958	1	V10328
38	19	54.3	1690	1	V10327
39	19	54.3	1714	1	V19790
40	19	54.3	2200	1	V44929
41	19	54.3	2200	1	V28288
42	19	54.3	2200	1	V32648
43	19	54.3	110000	1	X20248_03

ALIGNMENTS

44 19 54.3 5940 1 X23526 O. longistaminta X
45 19 54.3 8416 1 X23523 O. longistaminta X

```

RESULT 1
ID 059013 059013 standard; DNA; 38 BP.
AC 059013:
DT 12-APR-1994 (first entry)
DE Primer for p110 cDNA.
KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ss: amplification.
OS Synthetic.
PN W09321328-A.
PD 28-OCT-1993.
PF 13-APR-1993; G00761.
PR 13-APR-1992; GB-008135.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Dhond R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI Parker PJ, Volinia S, Waterfield MD;
PI WPI: 93-351738/44.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Example 1; Page 38; 146pp: English.
CC An SGBAF-1 cell line was established by transfection of bovine
CC adrenal cortex zona fasciculata cells with pSVneo. Total RNA was
CC isolated from the SGBAF-1 and a cDNA library constructed. The
CC unamplified library was plated on E. coli K12 PLF-F' and screened
CC with the labelled primer shown and the primer of Q59012. Hybridising
CC clones were sequenced.
CC See also Q51155-6, Q59012-23 and Q57522-3.
SQ Sequence 38 BP; 7 A; 6 C; 8 G; 11 T;

Query Match 94.3%; Score 33; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 2.8e-05;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATGATGCCARCTGTCATATGCAATTTGGCCA 35
DB 1 GATGATGCCARCTGTCATATGCAATTTGGCCA 35

RESULT 2
ID 057019 057019 standard; DNA; 38 BP.
AC 057019:
DT 31-AUG-1994 (first entry)
DE Platin 3-kinase probe #2.
KW 110 KD catalytic subunit; phosphatidyl inositol 3-kinase;
KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW Platin 3-kinase; assay; detection; cell growth; regulation; cancer;
KW blood vessel plaques; ss.
OS Synthetic.
PN W09403609-A.
PD 17-FEB-1994.
PF 05-AUG-1993; G01651.
PR 05-AUG-1992; GB-016654.
PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
PI WPI: 94-065687/08.
PT Eukaryotic cells transformed with mammalian phospholipid or
PT protein kinase DNA - useful in assays for compounds involved in
PT cell growth regulation and for treating cancers
PS Example 1; Page 14; 71pp: English.
CC The sequences given in 057018-19 are probes which were used in the
CC isolation of fusion sequences comprising the phosphatidyl inositol
CC (Pisins) 3-kinase under the regulatory control of the nmt promoter.

```

CC These sequences were transformed into Schizosaccharomyces pombe cells
 CC in an embodiment of the invention. In the presence of thiamine the
 CC promoter is inactive and the cells carrying the PKC plasmids grow as
 CC the parental strain. In the absence of thiamine the nmt promoter
 CC functions and the PKC is induced. PKC activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SO Sequence 38 BP; 7 A; 6 C; 8 G; 11 T;

Query Match 94.3%; Score 33; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 2,8e-05;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GATGATGCCACCTGTTTCATATGATGATTTGGCCA 35
 1 GATGATGCCACCTGTTTCATATGATGATTTGGCCA 35

RESULT 3
 ID 051155 standard; cDNA; 3207 BP.
 AC 051155:
 U1 12-APR-1994 (first entry)
 DE P110 cDNA.
 KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antaonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW plasmids; neurophil activity; 3-phosphorylated phosphoinositides;
 KW ss.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /note= "PI3- kinase p110"
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 DR P-PSDB: R43341.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Claim 7; Fig 9; 146pp; English.
 CC Southern blot analysis was performed using a bovine cDNA probe contg.
 CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
 CC from a cDNA library constructed from mRNA isolated from the human
 CC cell line K61a. Positive clones were sequenced to give the human
 CC PI3 kinase p110 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p110 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to
 CC provide a protein with PI3 kinase activity, and is useful for
 CC screening for (ant)agonists of PI3 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC prophylaxis or therapy. Platelet or neutrophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051156, 059012-23 and 057522-3.
 SO Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match 85.1%; Score 29.8; DB 1; Length 3207;
 Best Local Similarity 80.0%; Pred. No. 0.0018;
 Matches 28; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GATGATGCCACCTGTTTCATATGATGATTTGGCCA 35
 2773 GATGATGCCACCTGTTTCATATGATGATTTGGCCA 2807

RESULT 4
 ID 057012 standard; cDNA to mRNA; 3498 BP.
 AC 057012:
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase 110 kD catalytic subunit cDNA.
 KW 110 kD catalytic subunit; phosphatidy1 inositol 3-kinase;
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Bos taurus.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /product= p110

WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 DR P-PSDB: R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure; Fig 1; 71pp; English.
 CC This sequence encodes the 110 kD catalytic subunit of the phosphatidy1
 CC inositol (ptdins) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nmt promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the ptdins
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nmt promoter functions and the ptdins 3-kinase
 CC catalytic subunit is induced. Ptdins activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SO Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 85.1%; Score 29.8; DB 1; Length 3498;
 Best Local Similarity 80.0%; Pred. No. 0.0019;
 Matches 28; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GATGATGCCACCTGTTTCATATGATGATTTGGCCA 35
 2773 GATGATGCCACCTGTTTCATATGATGATTTGGCCA 2807

RESULT 5
 ID 051156 standard; cDNA; 3412 BP.
 AC 051156:
 DT 12-APR-1994 (first entry)
 DE Human p110 cDNA.
 KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antaonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neurophil activity; 3-phosphorylated phosphoinositides;
 KW ss.
 OS Human.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /note= "PI3- kinase p110"
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.


```

FT primer_bind /tag= a complement (34, .58)
FT primer_bind /tag= b /note= "primer for HDGF-2 recombinant prodn."
PD WO9639485-A1.
PN 12-DEC-1996.
PF 05-JUN-1995; U06731.
PR 05-JUN-1995; WO-U06731.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Kunsch CA, Rosen CA;
DR MPI; 97-043108/04.
DR P-PSDB; W09404.
PT DNA encoding human-derived growth factor polypeptide - useful to
PT promote wound healing as result of burns or ulcers
PS Claim 1, Fig 1A-C; 54pp; English.
CC A cDNA clone (747520) codes for human hepatoma-derived growth
CC factor-2 (HDGF-2) (W09404), a protein that can be used to
CC stimulate tissue repair and tissue growth. It was discovered in
CC a human umbilical vein endothelial tissue cDNA library, and may
CC also be obtained from heart, brain and skeletal muscle. HDGF-2
CC polynucleotides can be incorporated into vectors for prodn. of
CC recombinant HDGF-2 in host (e.g., E. coli, Sf9) cells. They can
CC also be used for the gene therapy of diseases related to HDGF-2
CC underexpression and for the design of diagnostic probes.
SQ Sequence 3202 BP; 817 A; 816 C; 819 G; 747 T;

Query Match 60.0%; Score 21; DB 1; Length 3202;
Best Local Similarity 68.6%; Pred. No. 8; 8;
Matches 24; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 GATGATGCCACTGTTTCATATGATGATTTGGCA 35
Ub 937 GATGCTGCCACTGTCCTGATGAGATGTCCA 903

RESULT 15
T80200
TD T80200 standard; cDNA: 6831 BP.
AC T80200;
UT 22-JUN-1998 (first entry)
DE Phosphatidyl inositol 3-kinase cdk cDNA.
KW Phosphatidyl inositol 3-kinase; signal transduction; cell cycle;
... autophagy; inflammatory joint disease; cell proliferation; cancer;
... psoriasis; restenosis; atherosclerosis; therapy; diagnosis; cpk; ss.
... Drosophila melanogaster.
FT key Location/Qualifiers
FT CDS 148..5778
FT /tag= a
PD WO9731650-A1.
PD 04-SEP-1997.
PF 12-FEB-1997; U02193.
PR 29-FEB-1996; US-609049.
PA (REGC ) UNIV CALIFORNIA.
PI Chen Y, Moiz L, Williams LT;
PI WPI; 97-448442/41.
PR P-PSDB; W38757.
PT New isolated phosphatidyl inositol-3 kinase polypeptide - used to
PT develop products for diagnosis and therapy, particularly for
PT proliferative disorders, e.g. inflammatory joint diseases, or cancer
PT Example 1; Fig 9; 77pp; English.
CC This cDNA sequence codes for cpk (see W38757), a Drosophila
CC polypeptide that belongs to a novel class of phosphatidyl inositol
CC 3-kinases that contain a C2 domain, are capable of phosphorylating
CC a D3 hydroxyl of an inositol ring in phosphatidyl inositol and
CC phosphatidyl inositol 4-phosphate, but not in phosphatidyl inositol
CC 4,5-bisphosphate, and which are involved in cell signalling cascades
CC that control e.g. cell cycle progression and intracellular protein
CC sorting. Short fragments of cpk cDNA were obtained from a
CC Drosophila cDNA library by PCR (see W38757). These short
CC fragments were used to screen the cDNA library to obtain larger
CC fragments, and missing 5' ends were obtained by RACE. A
CC recombinant host cell, transfected with a vector comprising a cpk

```

```

CC nucleic acid is claimed. Novel phosphatidyl inositol 3-kinases
CC can be used to screen for agonists/antagonists of activity and
CC in a claimed method of treating a disorder caused by dysregulation
CC of a growth factor activation signalling cascade. Antagonists
CC may reduce Ras activation allowing treatment of proliferative
CC disorders such as atherosclerosis, inflammatory joint disease,
CC psoriasis, restenosis following angioplasty, and cancer.
SQ Sequence 6831 BP; 1829 A; 1685 C; 1669 G; 1648 T;

Query Match 60.0%; Score 21; DB 1; Length 6831;
Best Local Similarity 72.4%; Pred. No. 11;
Matches 21; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 7 GGCCACTGTTTCATATGATGATTTGGCA 35
DB 4498 GGCTACTGTTTCACATGACTTGGCA 4526

Search completed: September 13, 1999, 15:59:29
Job time: 4796 sec

```

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Search time: September 13, 1999, 15:45:55 : Search time 1694.61 Seconds
(Without alignments)
40.740 Million cell updates/sec

Title: US-09-325-095-14

Perfect score: 35
Sequence: 1 GATGATGCCARCTGTCATCAATMGATTTGGCCA 35

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	28.2	80.6	453	45	A1324274	A1324274 mg72c02.y
2	25	71.4	476	29	A1174738	A1174738 mt12e11.r
3	24.8	70.9	328	21	R12466	R12466 yf56h07.r1
4	24.8	70.9	391	29	A1190747	A1190747 zp88405.r
5	22.4	64.0	479	25	N76563	N76563 yv40a03.r1
6	21.6	61.7	447	42	A115136	A115136 yd59i06.r
7	21.4	61.1	378	29	A1193439	A1193439 zr41b09.s
8	21.2	60.6	263	21	R18590	R18590 yf96b10.r1
9	21.2	60.6	509	21	R21085	R21085 y952f07.r1
10	21.2	60.6	330	22	R22689	R22689 y908e01.r1
11	21.2	60.6	574	37	AB009107	AB009107 AB009107
12	21	60.0	336	20	Z25224	Z25224 HSBABE042.S
13	20.8	59.4	497	29	A1172940	A1172940 mt01g12.r
14	20.8	59.4	446	29	A1175018	A1175018 ms82a03.r
15	20.8	59.4	517	29	A1175929	A1175929 ms92d03.r
16	20.8	59.4	494	30	AA243786	AA243786 zr67d12.r
17	20.8	59.4	452	31	AA287448	AA287448 z552d04.r
18	20.8	59.4	393	39	AA835841	AA835841 oc76d05.s
19	20.8	59.4	443	41	A1021241	A1021241 ub04b07.r
20	20.8	59.4	461	41	A1047071	A1047071 ub60h10.r
21	20.8	59.4	464	42	A1157114	A1157114 bd16d01.r
22	20.8	59.4	402	43	A1197569	A1197569 ud99a10.r
23	20.6	58.9	509	35	AA596554	AA596554 vms7d10.r
24	20.6	58.9	624	41	A1054648	A1054648 coa0001J
25	20.6	58.9	412	46	A1448463	A1448463 mg45d01.x
26	20.6	58.9	359	47	A1465317	A1465317 mu66h05.x
27	20.6	58.9	394	48	A1608079	A1608079 va08c12.y
28	20.6	58.9	328	48	A1661235	A1661235 va08c12.x
29	20.6	58.9	341	49	AV023366	AV023366 AV023366
30	20.4	58.3	338	21	T86707	T86707 yd86f07.s1
31	20.4	58.3	551	42	A1108767	A1108767 GH08006.5
32	20.4	58.3	509	42	A1108775	A1108775 GH08018.5
33	20.2	57.7	346	20	D24320	D24320 RICR1727A.R
34	20.2	57.7	373	22	H02035	H02035 yj38g08.s1
35	20.2	57.7	221	39	AA876400	AA876400 nv77h04.s
36	20.2	57.7	412	42	A1091644	A1091644 oc01g10.x
37	20.2	57.7	333	50	AV046640	AV046640 AV046640
38	20	57.1	329	48	A1600497	A1600497 486075F12
39	20	57.1	693	50	AU060839	AU060839 AU060839
40	20	57.1	591	51	A1734618	A1734618 606034A09
41	19.8	56.6	419	22	R26169	R26169 yb39g11.s1
42	19.8	56.6	352	40	AA932693	AA932693 oc074d02.s
43	19.6	56.0	139	29	AA168605	AA168605 nc30h01.r
44	19.6	56.0	624	29	AA901889	AA901889 NCP2ECT7
45	19.6	56.0	596	39	AA901891	AA901891 NCM4F4T3

ALIGNMENTS

RESULT 1
LOCUS A1324274 453 bp mRNA
DEFINITION mg72c02.y1 Striatogene mouse melanoma (#937312) Mus musculus cDNA
clone IMAGE:584258 5' similar to SW:P1A_MOUSE P42337
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
mRNA sequence.
ACCESSION A1324274


```

REFERENCE 1 (bases 1 to 328)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hiltman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Patson,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
* Trevaaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The Mashu-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Insert Size: 2060
High quality sequence stops: 256 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2060 Std Error: 0.00
Seq primer: M13Rpl
High quality sequence stop: 256.
Location/Qualifiers
1..328
/organism="Homo sapiens"
/db_xref="GDB:398791"
/db_xref="taxon:9606"
/clone="IMAGE:26444"
/clone_1lb="Soares infant brain IN1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lambda BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a NotB
I - 5'
AACGGAAGAATTCGGCGCCGACGAAATTTTATTTTATTTT 3'"];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lambda BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN
86 a 61 c 83 g 98 t
Query Match 70.9%; Score 24.8; DB 21; Length 328;
Best Local Similarity 76.7%; Pred. No. 1.7;
Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
CY 6 TGGCCARCTGTTTCATATGATTTGGCCA 35
|||||:||||:||||:||||:||||:
DB 110 TGGCCAGCTCTTCACATGACTTGACACA 139
|||||:||||:||||:||||:||||:
RESULT 4
LOCUS AA190747
DEFINITION zpr8605.r1 Stratogene Hela cell s3 937216 Homo sapiens cDNA clone
IMAGE:627273 5' similar to SW-PIB HUMAN P42338
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, BETA ISOFORM ;,
mRNA sequence.
ACCESSION AA190747
NID 91779133
VERSION AA190747.1 GI:1779133
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 391)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chipelli,B.,
Chissee,S., Dietrich,N., Dubuque,T., Favellio,A., Gish,W.,
Hawkins,M., Hiltman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,

```

TITLE
JOURNAL
MEDLINE

COMMENT

On May 18, 1995 this sequence version replaced gi:810974.

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LINT ; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 233.
Location/Qualifiers
1. 391
/organism="Homo sapiens"
/db_xref="GDB:5045971"
/db_xref="taxon:9606"
/clone="IMAGE:627273"
/clone_lib="Stratagene HeLa cell s3 937216"
/sex="female"
/dev_stage="HeLa S3 cell line"
/lab_host="SOLR (Kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3
epithelioid carcinoma cells grown to semi-confluency
without induction. Average insert size: 1.5 kb; Uni-ZAP
Vector. -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3'
adaptor sequence: 5' CTCGACGTTTCTTTTCTTTT 3' "

BASE COUNT
109 a 79 c 89 g 109 t 5 others

ORIGIN

Query Match
Best Local Similarity 70.9%; Score 24.8; DB 29; Length 391;
Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 6 TGGCCACGCTGTTTCATGATGATTTTGCCA 35
|||||:||||:||||:||||:|||||
Db 183 TGGCCAGCTCTTCACATGACTTGGACA 212

RESULT 5
N76563 LOCUS
DEFINITION
IMAGE:245164 5' similar to SW-PLIB HUMAN P42338
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, BETA ISOFORM ;
mRNA sequence.
N76563
91239141
N76563.1 GI:1239141
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 479)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiparelli, B.,
Chisose, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Maddis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rolling, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevaakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.

ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITTLE
JOURNAL
Genome Res. 6 (9), 807-828 (1996)
Generation and analysis of 280,000 human expressed sequence tags

MEDLINE 97044478
COMMENT On May 5, 1995 this sequence version replaced g1:798462

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu
This clone is available royalty-free through LNL: contact the
IMAG Consortium (info@image.lnl.gov) for further information
Insert length: 1532 Std Error: 0.00
Seq primer: reverse ET
High quality sequence stop: 366.

```

/organism="Homo sapiens"
/db_xref="GDB:3794410"
/db_xref="taxon:9606"
/map="21"
/clone_image="245164"
/clone_lib="Soares fetal liver spleen INTLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAACATCTCTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo."

```

Query Match	64.0%	Score 22.4	DB 25	length 479
Best Local Similarity	76.9%	Pred. No. 17		
Matches	20	Conservative	5	Mismatches 1
				Indels 0
				Gaps 0
QY	7	GGCCACRGTGTCATATGATTTTGG	32	
			
Db	359	GGCCAGCGTCTTCCACATGCACTTGG	384	

```

RESULT      6
AL155136
LOCUS
DEFINITION
A1155136          447 bp      mRNA
u699f06.r1 Soares mouse uterus NMU Mus musculus CDNA clone
IMAGE:1450211.5 similar to SW:P114.MOUSE P42337
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
mRNA sequence.
A1155136
A15513605
A1155136.1 GI:3683605
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 447)
Marra,M., Hillier,T., Allen,M., Bowles,M., Dierich,N., Dubouche,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,W., Tan,F., Underwood,K., Moore,B.,
Teisling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The Mashu-HMI Mouse EST Project
Unpublished (1996)
On Apr 18, 1995 this sequence version replaced gi:175514.

Contact: Marra M/Mouse EST project

```

Nashu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MG1:923527
Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 rev2. ET from Amersham
High quality sequence stop: 436.

```

FEATURES
source
location/Qualifiers
1..447
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1450211"
/clone_1fb="Soares mouse uterus NMPu"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; 1st strand cDNA was prepared from
pregnant mouse uterus, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Palma Bonaldo."
BASE COUNT
139 a 79 c 101 g 127 t 1 others
ORIGIN
Query Match 61.7%; Score 21.6; DB 42; Length 447;
Best Local Similarity 70.0%; Pred. No. 36;
Matches 21; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

Qy	6	TGGCCARCTGTTTCACATATMGAAVTTGGCCA	35
Db	21	TTGAAGCTGTTTCATATAGATTTGGGCA	50

	RESULT	7		
LOCUS	AAI193439			
DEFINITION	AI193439	378 bp	mRNA	EST
	zr4ID09.s1 Soares_NIHMPu_S1 Homo sapiens cDNA clone IMAGE:665945			19-MAY-1997
ACCESSION	AAI193439			
NID	G1782867			
VERSION	AAI193439.1	GI:1782867		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 378)			
AUTHORS	Hillier,L., Lennon,G., Becker,M., Donald,M.F., Chapell,I.B., Chisoe,S., Dietrich,N., Dubugue,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L. Roefling,T., Schellendberg,K., Soares,M.B., Tan,F., Thierry-Mieg,D. Trevaaks,E., Underwood,K., Woldmann,P., Waterston,R., Wilson, and Marra,M.			
TITLE	Generation and analysis of 280,000 human expressed sequence tags			
JOURNAL	Genome Res.	6 (9),	807-828	(1996)
MEDLINE	97044478			
COMMENT	On Sep 1, 1995 this sequence version replaced.			

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810

source

1. 509

/organism="Homo sapiens"

/db_xref="GDB:408627"

/db_xref="taxon:9606"

/map="9"

/clone="IMAGE:36126"

/clone_lib="Soares infant brain INIB"

/sex="female"

/dev_stage="73 days post natal"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: whole brain; Vector: lacmid BA; Site.1: Not I; Site.2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5];

ACTGAGAGAAATTCGCGCCGACAGAAATTTTCTTTTCTTTT 3';

double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the lacmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 134 a 104 c 116 g 143 t 12 others

ORIGIN

Query Match 60.6%; Score 21.2; DB 21; Length 509;

Best Local Similarity 76.9%; Pred. No. 53;

Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 7 GGCCACATGTTTCATATGACTTTGG 32

||||| |||||:||||:|||||

UP 87 GGCCACATGTTTCATATGACTTTGG 112

RESULT 10

LOCUS R22699 330 bp mRNA EST 19-APR-1995

DEFINITION YG08e01.r1 Soares infant brain INIB Homo sapiens cDNA clone

IMAGE:31318 5' similar to SP:P100_BOVIN P32871 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT ;, mRNA sequence.

ACCESSION R22699

NID 9777456

VERSION R22699.1 GI:777456

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS 1 (bases 1 to 330)

Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marre,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT

*Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

Insert Size: 1762

High quality sequence stops: 271 Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 1762 Std Error: 0.00

Seq primer: M13p1

High quality sequence stop: 271.

FEATURES

Location/Qualifiers

1..330

/organism="Homo sapiens"

/db_xref="GDB:403655"

/db_xref="taxon:9606"

source

/clone="IMAGE:31318"

/clone_lib="Soares infant brain INIB"

/sex="female"

/dev_stage="73 days post natal"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: whole brain; Vector: lacmid BA; Site.1: Not I; Site.2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5];

ACTGAGAGAAATTCGCGCCGACAGAAATTTTCTTTTCTTTT 3';

double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the lacmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 92 a 62 c 77 g 99 t

ORIGIN

Query Match 60.6%; Score 21.2; DB 22; Length 330;

Best Local Similarity 76.9%; Pred. No. 50;

Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 7 GGCCACATGTTTCATATGACTTTGG 32

||||| |||||:||||:|||||

DB 87 GGCCACATGTTTCATATGACTTTGG 112

RESULT 11

LOCUS AB009107 574 bp mRNA EST 05-DEC-1997

DEFINITION AB009107 Chlamydomonas W80 lambda ZAP II Chlamydomonas sp. cDNA

similar to phosphatidylinositol 3-kinase, mRNA sequence.

ACCESSION AB009107

NID 92662210

VERSION AB009107.1 GI:2662210

KEYWORDS EST.

SOURCE Chlamydomonas sp.

ORGANISM Chlamydomonas sp.

REFERENCE Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

AUTHORS 1 (bases 1 to 574)

Kanaboshi,H., Ikeda,K. and Miyasaka,H.

TITLE Isolation of several anti-stress genes from the halotolerant green alga Chlamydomonas by a simple functional expression screening in E. coli

JOURNAL Unpublished (1998)

COMMENT On Sep 19, 1997 this sequence version replaced gi:1520267.

CONTACT: Hitoshi Miyasaka

Technical Research Center

The Kansai Electric Power Co.

Nakoji 3-Chome 11-20, Amagasaki, Hyogo 661, Japan

Tel: 81-6-494-9840

Fax: 81-6-494-9728

Email: hmiya@emb.infoweb.or.jp.

FEATURES

Location/Qualifiers

1..574

/organism="Chlamydomonas sp."

/strain="W80"

/db_xref="taxon:3056"

/clone_lib="Chlamydomonas W80 lambda ZAP II"

/note="Vector: lambda ZAP II; The cDNA clone was isolated from the halotolerant green alga Chlamydomonas W80 by a functional expression screening in E. coli cells. The principle of the screening method was based on the acquisition of stress tolerance of the bacterial cells carrying the cDNA."

BASE COUNT 96 a 197 c 184 g 97 t

ORIGIN

Query Match 60.6%; Score 21.2; DB 37; Length 574;

Best Local Similarity 71.9%; Pred. No. 54;

Matches	23;	Conservative	3;	Mismatches	6;	Indels	0;	Gaps	0;
Qy	4	GATGGCCARCTGTTTCATATMGATTTTGCGCA	35						
Db	458	GACGGCCGCTCTTCACATCGACTTGGCTA	489						
RESULT	12								
LOCUS	225224/c								
DEFINITION	225224	336 bp	mRNA	EST	02-AUG-1993				
	HSNA6042	STRATIGENE	Human skeletal muscle cDNA library,	cat.					
	#936215.	Homo sapiens	CDNA clone A6E04,	mRNA sequence.					
ACCESSION	225224								
NID	9395863								
VERSION	225224.1	GI:395863							
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;								
	Eutheria; Primates; Catarrhini; Hominoidea; Homo.								
REFERENCE	1 (bases 1 to 336)								
AUTHORS	Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,								
	Devignes,M.D., Duprat,S., Hougate,R., Juneau,M.N., Lamy,B.,								
	Lozano,F., Mitchell,H., Marigoe-Samson,R., Pletu,G., Pouliot,Y.,								
	Sébastia,I., Kabackitch,C. and Tessier,A.								
TITLE	IMAGE: molecular integration of the analysis of the human genome								
	and its expression								
JOURNAL	C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)								
MEDLINE	95277534								
COMMENT									

```

Query Match          60.0%: Score 21: DB 20: Length 336;
Best Local Similarity 68.6%: Pred. NO. 61;
Matches 24: Conservative 3; Mismatches 8; Indels 0: Gaps 0
UY 1 GATGATGGCCARCCTGTTCAVATGAGATTGGCCA 35
    ||| |||||:||||:| ||| || |||
Db 201 GATGCTGGCCAACTGTCCCTGATTGAGATGCCA 167

RESULT 13
LOCUS AA172940/c
DEFINITION mt01912.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:619846
5', mRNA sequence.
ACCESSION AA172940
G1751992
NID AA172940.1 GI:1751992

FEATURES
Source
1..336
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="A6E04"
/clone_1ib="STRATAGENE Human skeletal muscle cDNA library
cat. #936215."
/sex="female"
/tissue_type="skeletal muscle"
/dev_stage="19 years"
/note="Organ: leg muscle; Vector: Lambda ZAPII; Tissue
from female, 19 years old, normal leg muscle. Cloning
vector is Lambda ZAPII, in vivo excision from Lambda ZAPII
to pBluescript SK(+). Genexpress library reference 1s B."
BASE COUNT 79 a 89 c 88 g 76 t 4 others
C1C1C1C

```

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 497)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Gessel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellensberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B.,
Teisberg,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On May 5, 1995 this sequence version replaced gi:7968375.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:380670
Seq primer: -28M13 rev2 from Amer sham
High quality sequence stop: 496.
Location/Qualifiers
1..497

	/organism="Mus musculus"
	/strain="C57BL/6J"
	/db_xref="taxon:10090"
	/map="21"
	/clone="IMAGE:619846"
	/clone_lib="Soares mouse 3BMDs"
	/sex="male"
	/tissue_type="Spleen"
	/dev_stage="4 weeks"
	/lab_host="DH10B"
	/note=Vector: pT73D-Pac (Pharmacia) with a modified
	polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
	was primed with a Not I - oligo(dT) primer [5'
	TGGTACCAATCTGAAGTGSGAGCGCCGCTGTATTTTTTTTTTTTTTT
	3']; double-stranded cDNA was ligated to Eco RI adaptors
	(Pharmacia), digested with Not I and cloned into the Not I
	and Eco RI sites of the modified pT73 vector. RNA
	provided by Dr. Bertrand Jordan. Library went through
	three rounds of normalization, and was constructed by
	Bento Soares and M.Falina Bonaldo."
BASE COUNT	159 a 62 c 97 g 179 t
ORIGIN	
Query Match	59.4% Score 20.8 DB 29 Length 497;
Best Local Similarity	68.8%; Pred. No. 77;
Matches 22: Conservative	4; Mismatches 6; Indels 0; Gaps 0;
DQ	2 ATGATGCCARCCTGTYCAVATWGAAYTTTCGC 33 : : Db 311 ATACAGCGCAACTGTACCACATAGATTGGC 280
RESULT 14	
AAl175018/c	
LOCUS	AAl175018 446 bp mRNA EST 16-FEB-1997
DEFINITION	ms82a03.r1 Soares mouse 3BMDs Mus musculus cDNA clone IMAGE:618028
ACCESSION	AAl175018
NID	G1756158
VERSION	AAl175018.1 GI:1756158
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE AUTHORS
1 (bases 1 to 446)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Matta,M., Hille,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
The Washu-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1290366.

TITLE
JOURNAL
COMMENT

FEATURES
SOURCE
1. 446
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="901A02:1;1p36.13-1p36.33"
/clone="IMAGE:618028"
/clone_1b="Soares mouse 3NDMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15', TGTTACCATCTGTGAGTGGAGCGCGCGCGCTGTTTTTTTTTTTTTTT 3') double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRTT3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
ORIGIN
148 a 63 c 79 g 156 t

Query Match 59.4%; Score 20.8; DB 29; Length 446;
Best Local Similarity 68.8%; Pred. NO. 76;
Matches 22; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
0Y 2 ATGATGGCCARCTGTTCAATAGATTTGGC 33
||| |||||:||||:||||:| |||||
Db 202 ATACAGCCAACTGTACCACATAGATTGGC 171

RESULT 15
LOCUS 517 bp mRNA EST 16-FEB-1997
DEFINITION ms92d03.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:619013
5', mRNA sequence.
ACCESSION AA175929
NID 61757049
VERSION AA175929.1 GI:1757049
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 517)

TITLE
The WashU-HHMI Mouse EST Project
JOURNAL COMMENT
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced g1.1324664.

AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubouque,T., Geisels,S., Kincade,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

FEATURES
SOURCE
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (InfoImage.lnl.gov) for further information.
GI:379837
Seq primer: -28M13 rev2 from Amesham
High quality sequence stop: 497.
Location/Qualifiers
1..517
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="4 q25-q35"
/clone="IMAGE:619013"
/clone_lib="Soares mouse 3mbMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5', TGTTACCAATCTGAAGTGGGAGCGCCGCCGTCTTTTTTTTTTTTTTTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
164 a 67 c 92 g 194 t

ORIGIN

Query Match 59.4%; Score 20.8; DB 29; Length 517;
Best Local Similarity 68.8%; Pred. No. 78;
Matches 22; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 ATGATGGCCARCTGTTTCAYATWGAATTGGC 33
||| ||||| :||| :||| :|||
Db 360 ATAACGACCACTGTACCACATAGATATTGGC 329

Search completed: September 13, 1999, 15:45:55
Job time: 4043 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:15 ; Search time 539.84 Seconds

(without alignments)
135,498 Million cell updates/sec

Title: US-09-325-095-15

Perfect score: 23

Sequence: 1 AATTGACACACTGGCATGCCGAT 23

Scoring table: IDENTITY_NDC

Searched: 679419 seqs, 1590154680 residues

Database:

GenEmbl:.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pi1.*
8: gb_pi2.*
9: gb_pi3.*
10: gb_pi4.*
11: gb_pi5.*
12: gb_pi6.*
13: gb_pi7.*
14: gb_pi8.*
15: gb_pi9.*
16: gb_pi10.*
17: gb_pi11.*
18: gb_pi12.*
19: gb_pi13.*
20: gb_pi14.*
21: gb_pi15.*
22: gb_pi16.*
23: gb_pi17.*
24: gb_pi18.*
25: gb_pi19.*
26: gb_pi20.*
27: gb_pi21.*
28: gb_pi22.*
29: gb_pi23.*
30: gb_pi24.*
31: gb_pi25.*
32: gb_pi26.*
33: gb_pi27.*
34: gb_pi28.*
35: gb_pi29.*
36: gb_pi30.*
37: gb_pi31.*
38: gb_pi32.*
39: gb_pi33.*
40: gb_pi34.*
41: gb_pi35.*
42: gb_pi36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	100.0	3207	3	BOVPHOS3KN
2	100.0	3207	3	BOVPHOS3KN
3	100.0	3207	3	BOVPHOS3KN
4	100.0	3207	3	BOVPHOS3KN
5	100.0	3207	3	BOVPHOS3KN
6	100.0	3207	3	BOVPHOS3KN
7	100.0	3207	3	BOVPHOS3KN
8	100.0	3207	3	BOVPHOS3KN
9	100.0	3207	3	BOVPHOS3KN
10	100.0	3207	3	BOVPHOS3KN
11	100.0	3207	3	BOVPHOS3KN
12	100.0	3207	3	BOVPHOS3KN
13	100.0	3207	3	BOVPHOS3KN
14	100.0	3207	3	BOVPHOS3KN
15	100.0	3207	3	BOVPHOS3KN
16	100.0	3207	3	BOVPHOS3KN
17	100.0	3207	3	BOVPHOS3KN
18	100.0	3207	3	BOVPHOS3KN
19	100.0	3207	3	BOVPHOS3KN
20	100.0	3207	3	BOVPHOS3KN
21	100.0	3207	3	BOVPHOS3KN
22	100.0	3207	3	BOVPHOS3KN
23	100.0	3207	3	BOVPHOS3KN
24	100.0	3207	3	BOVPHOS3KN
25	100.0	3207	3	BOVPHOS3KN
26	100.0	3207	3	BOVPHOS3KN
27	100.0	3207	3	BOVPHOS3KN
28	100.0	3207	3	BOVPHOS3KN
29	100.0	3207	3	BOVPHOS3KN
30	100.0	3207	3	BOVPHOS3KN
31	100.0	3207	3	BOVPHOS3KN
32	100.0	3207	3	BOVPHOS3KN
33	100.0	3207	3	BOVPHOS3KN
34	100.0	3207	3	BOVPHOS3KN
35	100.0	3207	3	BOVPHOS3KN
36	100.0	3207	3	BOVPHOS3KN
37	100.0	3207	3	BOVPHOS3KN
38	100.0	3207	3	BOVPHOS3KN
39	100.0	3207	3	BOVPHOS3KN
40	100.0	3207	3	BOVPHOS3KN
41	100.0	3207	3	BOVPHOS3KN
42	100.0	3207	3	BOVPHOS3KN
43	100.0	3207	3	BOVPHOS3KN
44	100.0	3207	3	BOVPHOS3KN
45	100.0	3207	3	BOVPHOS3KN

C	2	23	100.0	3498	5	A37232	A37232 Sequence 1
C	3	23	100.0	223	5	A37243	A37243 Sequence 12
C	4	23	100.0	3207	10	HSU79143	U79143 Human phosph
C	5	21.4	93.0	3424	10	HSPI3K	229090 H.sapiens m
C	6	21.4	93.0	3207	12	MMU03279	U03279 Mus musculus
C	7	18.2	79.1	640	14	G30932	G30932 pig STS CP-
C	8	17.2	74.8	65020	11	AC003691	AC003691 Human Chr
C	9	17.2	74.8	5085	12	MMANREC	211974 M.musculus
C	10	17.2	74.8	190000	35	AC000642	AC000642 Homo sapi
C	11	17.2	74.8	39838	42	AC007766	AC007766 Homo sapi
C	12	16.8	73.0	6067	1	CDCCYNAB	M34459 C.saccharol
C	13	16.8	73.0	18483	2	AF005383	AF005383 Caldicell
C	14	16.8	73.0	2062	5	184210	184210 Sequence 8
C	15	16.8	73.0	164	5	184217	184217 Sequence 19
C	16	16.8	73.0	95214	8	ATAC06526	AC006526 Arabidops
C	17	16.8	73.0	110608	9	HS283E3	AL031482 Human DNA
C	18	16.8	73.0	107418	9	HS51E13	AL022163 Human DNA
C	19	16.8	73.0	5586	12	D67015	D67015 Mouse mRNA
C	20	16.8	73.0	3262	17	PPSKS1A	L39364 Porcine res
C	21	16.8	73.0	3292	17	PPSG1A	L39369 Porcine res
C	22	16.8	73.0	2050	17	PRU03040	U03040 Porcine rep
C	23	16.8	73.0	167439	35	AC007239	AC007239 Drosophila
C	24	16.6	72.2	9117	1	PEMETH	L29642 Pseudomonas
C	25	16.6	72.2	2179	3	CAPR4	AF079548 Ovis arie
C	26	16.6	72.2	10763	5	AR008973	AR008973 Sequence
C	27	16.6	72.2	5559	5	166429	166429 Sequence 3
C	28	16.6	72.2	5559	5	173354	173354 Sequence 3
C	29	16.6	72.2	5559	5	173361	173361 Sequence 3
C	30	16.6	72.2	5559	5	181321	181321 Sequence 3
C	31	16.6	72.2	104364	8	ATAC006919	AC006919 Arabidops
C	32	16.6	72.2	114929	9	AP000050	AP000050 Homo sapi
C	33	16.6	72.2	151516	9	AP000117	AP000117 Homo sapi
C	34	16.6	72.2	100000	9	AP000193	AP000193 Homo sapi
C	35	16.6	72.2	78539	9	AP000313	AP000313 Homo sapi
C	36	16.6	72.2	167514	9	HS242N11	AL023655 Human DNA
C	37	16.6	72.2	154902	11	AC005799	AC005799 Homo sapi
C	38	16.6	72.2	369	11	HS1SN3	AF064247 Homo sapi
C	39	16.6	72.2	5366	17	HAPVX	X02449 Hamster pap
C	40	16.6	72.2	5366	17	PPCGGAA	M26281 Hamster pap
C	41	16.6	72.2	133572	35	AC006965	AC006965 Homo sapi
C	42	16.6	72.2	39478	36	CEF3E1	281068 Caenorhabdi
C	43	16.6	72.2	34190	36	CEK4G11	278544 Caenorhabdi
C	44	16.6	72.2	91019	37	AC001658	AC001658 Drosophila
C	45	16.6	72.2	175775	42	AC006947	AC006947 Homo sapi

ALIGNMENTS

RESULT 1
BOVPHOS3KN/C
LOCUS
DEFINITION Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete
CDS
ACCESSION M93252
NTD 9163519
VERSION M93252.1 GI:163519
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE Bos taurus cDNA to mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 3207)
Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R.,
Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F.,
Hsuan, J.J., Courtneidge, S.A., Parker, P.J. and Waterfield, M.D.
Phosphatidylinositol 3-kinase: Structure and expression of the 110
kd catalytic subunit
JOURNAL MEDLINE
Cell 70, 419-429 (1992)
92354059
FEATURES
location/Qualifiers
1..3207

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS 1 (bases 1 to 3207)
TITLE Olfiff, A. and Heimbrock, D.C.
Catalytic Activity of the p110-alpha Subunit of Human Phosphoinositide 3'-Hydroxylase is Required for Signal Transduction
JOURNAL Blood. Med. Chem. (1996) In press
REFERENCE 2 (bases 1 to 3207)
AUTHORS Stirdivant, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledder, L.M., Olfiff, A. and Heimbrock, D.C.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs, Sumneytown Pike, West Point, PA 19486, USA
FEATURES
SOURCE
CDS
1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3207
/function="phospholipid kinase"
/codon_start=1
/product="phosphoinositide 3'-hydroxylase p110-alpha subunit"
/protein_id="AAB39753.1"
/db_xref="PID:q1763626"
/db_xref="GI:1763626"
/translation="MPPRSSGELMGHILMPRIIVLVECLLPNGMIVTECREATLV
IKHLEFKARKYPLHQLDSESYFVSVDGEAREEFEDTRLCIDRLRPLKVI
EPVGNREKILNREIGFALGMPVCEFMVNDPEVDPEFRLNLVCKEAVDRLDSPH
SRATVYVPVPHVSESPKPHLYNKLNGQIIVIVIVSPNNDKOKYTLKLNHCVE
QVIAEARKTRSMILSEQLKLVLEQGYILKVCDDYFLEKPLSOYKTRSC
IMGRMNLMMAKESLYSOLPMDCFMPSRSRISTATPMNGETSRKSLMVTNRL
RIKILCATYVNLNLDIKIVRGYIGSGPLCDNVTQVPCSNPMNMLVNDIY
IPDLPBARCLCISYKGRGAKBEHCPPLMNGINLEFDITDVLVSGMALNMPVPH
GLEDLNPGVTSNPKETPCLEDFDSESVVKFPMDSVIEEHANVSREGEFSY
SHAOLSNRLANDNELRENDKQLAISTRDPLSEIYDEKDFLMSHRVYCYTIEILP
KLLSVKNSNDEVAQWYCLVNDPPIKEQAMELLDCNYPDPVGRGAVCLERKYL
DRLSOYLIOVLYKYEQYLDNLVRLKALTNORIGHEFFMHLKSEHNNTVSO
RPPMDALOGFLSPNPAHQGNRLRECRINSKRPRLMNPENPDIINSLLFQNSH
IIFKNGDLRODMTLQIIRIMENIMONQGLDMLPYGCLSIDCVGLLIVVNSHT
IMJOQCGKLGALQFNSHTLHOMLKDKNKEIYDAIDFTFSCACGCVATFLLGIG
DRHNSNINVKDQGLFHDGFHDKKKKRGYRERPEVLTODFLVLSKGOECT
KTRFEERQEMCYAYLAIRQHANLFTINLSMMLGSGMPELQSFDDIAYIRKTLALDK
TEOALTEFMKQNDAHGCGTTKMDVTFHTIKOHALN"

BASE COUNT 1043 a 586 c 670 g 908 t
ORIGIN

Query Match 100.0%; Score 23; DB 10; Length 3207;
Best Local Similarity 100.0%; Pred. No. 0.08; Mismatches 0; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTCACACTGCATGCCGAT 23
|||||
Db 383 AATTCACACTGCATGCCGAT 361

RESULT 5
HSPH13K 3424 bp mRNA PRI 24-AUG-1995
LOCUS H.sapiens mRNA for phosphatidylinositol 3-kinase.
DEFINITION
ACCESSION Z29090
NID 9472990
VERSION 229090.1 GI:472990
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

REFERENCE
AUTHORS 1 (bases 1 to 1068)
TITLE Volinia, S., Hiles, I., Ormondo, E., Mizetic, D., Antonacci, R., Rocchi, M. and Waterfield, M.D.
Molecular cloning, cDNA sequence, and chromosomal localization of the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene
JOURNAL Genomics 24 (3), 472-477 (1994)
REFERENCE 2 (bases 1 to 3424)
AUTHORS Volinia, S.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig Institute for Cancer, Research, 91 Riding House Street, London, W1P 8BT, UK
FEATURES
SOURCE
CDS
1..3424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="K562"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CA82333.1"
/db_xref="PID:q472991"
/db_xref="GI:472991"
/db_xref="SWISS-PROT:P42336"
/translation="MPPRSSGELMGHILMPRIIVLVECLLPNGMIVTECREATLV
IKHLEFKARKYPLHQLDSESYFVSVDGEAREEFEDTRLCIDRLRPLKVI
EPVGNREKILNREIGFALGMPVCEFMVNDPEVDPEFRLNLVCKEAVDRLDSPH
SRATVYVPVPHVSESPKPHLYNKLNGQIIVIVIVSPNNDKOKYTLKLNHCVE
QVIAEARKTRSMILSEQLKLVLEQGYILKVCDDYFLEKPLSOYKTRSC
IMGRMNLMMAKESLYSOLPMDCFMPSRSRISTATPMNGETSRKSLMVTNRL
RIKILCATYVNLNLDIKIVRGYIGSGPLCDNVTQVPCSNPMNMLVNDIY
IPDLPBARCLCISYKGRGAKBEHCPPLMNGINLEFDITDVLVSGMALNMPVPH
GLEDLNPGVTSNPKETPCLEDFDSESVVKFPMDSVIEEHANVSREGEFSY
SHAOLSNRLANDNELRENDKQLAISTRDPLSEIYDEKDFLMSHRVYCYTIEILP
KLLSVKNSNDEVAQWYCLVNDPPIKEQAMELLDCNYPDPVGRGAVCLERKYL
DRLSOYLIOVLYKYEQYLDNLVRLKALTNORIGHEFFMHLKSEHNNTVSO
RPPMDALOGFLSPNPAHQGNRLRECRINSKRPRLMNPENPDIINSLLFQNSH
IIFKNGDLRODMTLQIIRIMENIMONQGLDMLPYGCLSIDCVGLLIVVNSHT
IMJOQCGKLGALQFNSHTLHOMLKDKNKEIYDAIDFTFSCACGCVATFLLGIG
DRHNSNINVKDQGLFHDGFHDKKKKRGYRERPEVLTODFLVLSKGOECT
KTRFEERQEMCYAYLAIRQHANLFTINLSMMLGSGMPELQSFDDIAYIRKTLALDK
TEOALTEFMKQNDAHGCGTTKMDVTFHTIKOHALN"

BASE COUNT 1134 a 618 c 709 g 963 t
ORIGIN

Query Match 93.0%; Score 21.4; DB 10; Length 3424;
Best Local Similarity 95.7%; Pred. No. 0.53; Mismatches 1; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AATTCACACTGCATGCCGAT 23
|||||
Db 395 AATTCACACTGCATGCCGAT 373

RESULT 6
MKU03279/c 3207 bp mRNA ROD 25-MAY-1994
LOCUS Mus musculus Balb/c phosphatidylinositol 3-kinase 110 kDa subunit
DEFINITION
ACCESSION U03279
NID 9414994
VERSION U03279.1 GI:414994
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3207)

AUTHORS Klippel, A., Escobedo, J.A., Hirano, M. and Williams, L.T.
TITLE The interaction of small domains between the subunits of phosphatidylinositol 3-kinase determines enzyme activity
JOURNAL Mol. Cell. Biol. 14, 2675-2685 (1994)
MEDLINE 94187738
REFERENCE 2 (bases 1 to 3207)
AUTHORS Klippel-Giese, A.
TITLE Direct Substitution
JOURNAL Submitted (05-NOV-1993) Anke Klippel-Giese, CVRI and Dalich Research Center, University of California San Francisco, 505 Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
FEATURES
 source
 1..3207
 /organism="Mus musculus"
 /strain="Balb/c"
 /db_xref="taxon:10090"
 1..3207
 /standard_name="p110 protein"
 /codon_start=1
 /product="phosphatidylinositol 3-kinase 110 kDa subunit"
 /protein_id="AA18334.1"
 /db_xref="PID:9414995"
 /db_xref="GI:414995"
 /translation="MPPRSGSELMGIIHMPRIIVCELLPQMGIVTECLREATIVT IKHELFRERKYPHLOLLODEPTSYIFSVTOAEEREFEDETRLCDDLRLPPLKVI EFGVNRREKILREIFGIVGMPVCEFDNPKDPEYDFRNINLVCKEADVLNDSPH SRAMVYIPNVSSPELPKRIITKLDKGIIVAVIVSFPNDKQITLKHNDVPE OVAIAIRKTRSMILSSBOLKLVLEYGKYLKVCQDEFFLEYSQYKIRISC IMGRPNIMLMAKESLYSQLIDSEFTMPSRSRISTAPYNGESTKSLVYNISAL RIKLCATVYNNIRIDIKIVRTGYHGEPLCDVNTORVPCSNPNNENLVNDIY IDPLRLARLCSICSGKRGAKHECHLANGNINLFPYTDIVSGKALNMLPVPY GLEDLNPGVGSNPKRPELPELPESSVYKPRDMSVIEHANSVSEADGSPS SHTGJSNRLARNELEBNDKEJRLICTDPISEITGEKDLWSRHRICVTPPELP KLLLSVKNMSRDEVAOMYCLVDWPIKEQKAMELDCNYPDPVASEPVRQLELYT DKLISQYLIQIVQVLYEQYLDNLVFLKRALNQRIGHEFWMLKSEMKNTVQ REGLLSEYCRACGMIKLNQVEMEKILNLDLQEKDEQKQVMKFLVQMR ODFMDALOGFLSPNPAHQLNLRLECRIMSARPKPLMKNENDIMSELFOVNE IIFKGGDDLRQDMITQITRIMENIMONGDLPMPYCLSTGCVGLIEVRSHT IMQIOCKGLKRALQFNSTHTLQWLNDRKGETIYDAIDLFTRSCAGTCVATFTIGIG DRHNSNINVKDQGLHDEGFHDLHKKRKFYKREKRVFVLDPLIVISGAGQET KTRFEREOMCYKAVLAIRQHANLEINFSMLSGMPELOSFDIAYIRKTLALDK TQEALLEYTKQMDNDAHSGMTTKMDIMFTIKQHLN"

CDS
 /product="phosphatidylinositol 3-kinase 110 kDa subunit"
 /protein_id="AA18334.1"
 /db_xref="PID:9414995"
 /db_xref="GI:414995"
 /translation="MPPRSGSELMGIIHMPRIIVCELLPQMGIVTECLREATIVT IKHELFRERKYPHLOLLODEPTSYIFSVTOAEEREFEDETRLCDDLRLPPLKVI EFGVNRREKILREIFGIVGMPVCEFDNPKDPEYDFRNINLVCKEADVLNDSPH SRAMVYIPNVSSPELPKRIITKLDKGIIVAVIVSFPNDKQITLKHNDVPE OVAIAIRKTRSMILSSBOLKLVLEYGKYLKVCQDEFFLEYSQYKIRISC IMGRPNIMLMAKESLYSQLIDSEFTMPSRSRISTAPYNGESTKSLVYNISAL RIKLCATVYNNIRIDIKIVRTGYHGEPLCDVNTORVPCSNPNNENLVNDIY IDPLRLARLCSICSGKRGAKHECHLANGNINLFPYTDIVSGKALNMLPVPY GLEDLNPGVGSNPKRPELPELPESSVYKPRDMSVIEHANSVSEADGSPS SHTGJSNRLARNELEBNDKEJRLICTDPISEITGEKDLWSRHRICVTPPELP KLLLSVKNMSRDEVAOMYCLVDWPIKEQKAMELDCNYPDPVASEPVRQLELYT DKLISQYLIQIVQVLYEQYLDNLVFLKRALNQRIGHEFWMLKSEMKNTVQ REGLLSEYCRACGMIKLNQVEMEKILNLDLQEKDEQKQVMKFLVQMR ODFMDALOGFLSPNPAHQLNLRLECRIMSARPKPLMKNENDIMSELFOVNE IIFKGGDDLRQDMITQITRIMENIMONGDLPMPYCLSTGCVGLIEVRSHT IMQIOCKGLKRALQFNSTHTLQWLNDRKGETIYDAIDLFTRSCAGTCVATFTIGIG DRHNSNINVKDQGLHDEGFHDLHKKRKFYKREKRVFVLDPLIVISGAGQET KTRFEREOMCYKAVLAIRQHANLEINFSMLSGMPELOSFDIAYIRKTLALDK TQEALLEYTKQMDNDAHSGMTTKMDIMFTIKQHLN"

BASE COUNT 958 a 693 c 734 g 822 t
ORIGIN

Query Match 93.0%; Score 21.4; DB 12; Length 3207;
Best Local Similarity 95.7%; Pred. No. 0.53;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AATTCACACACTGGCATGCCGAT 23
 ||||||||||||||||||||
DB 383 AATTCACACACTGGCATGCCAAT 361

RESULT 7
LOCUS G30932 640 bp DNA STS 15-NOV-1996
DEFINITION pig STS CP-000R, sequence tagged site.
ACCESSION G30932
NID 91669818
VERSION G30932.1 GI:1669818
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE pig strain-Melishan Pig genomic DNA was prepared by standard procedure.
ORGANISM Sus scrofa
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
AUTHORS 1 (bases 1 to 640)
JOURNAL Comparative gene mapping between human and pigs
COMMENT Unpublished (1996)

Contact: Sun, H. S.
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150
 Tel: 515-294-4209
 Fax: 515-294-2401
 Email: hssun@iastate.edu

Primer A: CTAGTCCTGTCATTGGC
Primer B: TCTTTGGGAGACGTCATTC
STS size: 640
PCR Profile:

Presoak: 95 degree C for 2 minutes
Denaturation: 95 degree C for 0.5 minute
Annealing: 58 degree C for 1 minute
Polymerization: 72 degree C for 1 minutes
PCR cycles: 32
Thermal Cycler: MJ Research

Protocol:
 Template: 30-100 ng
 Primers: 0.3 uM
 dNTPs: each 200 uM
 Tag Polymerase: 0.033 units/ul
 Total Vol: 15 ul

Buffer:

MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 8.3.

FEATURES
 source
 1..640
 Location/Qualifiers
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /strain="Melishan"

STS
 >1..<640
 /gene="CP"
 <1..>640
 /gene="CP"
 /note="ceruloplasmin"

BASE COUNT 186 a 126 c 142 g 186 t
ORIGIN

Query Match 79.1%; Score 18.2; DB 14; Length 640;
Best Local Similarity 87.0%; Pred. No. 27;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AATTCACACACTGGCATGCCGAT 23
 ||||||||||||||||||||
DB 292 AATTCACACACTGGCATCTAAT 314

RESULT 8
LOCUS AC003691 65020 bp DNA PRI 29-OCT-1998
DEFINITION Human Chromosome 11q23 PAC clone pD754e13, complete sequence.
ACCESSION AC003691
NID 93808083
VERSION AC003691.1 GI:3808083
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basil, M., Buettner, T., Bumsister, R., Card, P., Desaliboat, F., Dunn, J., English, C., Ethridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G., Grant, O., Hamner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N., Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S., Schageman, J., Schultz, R.A., Stimson, S., Syed, M. and Ward, T.
TITLE HTGS Submission

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 65020)
AUTHORS Evans,G.A., Athanasiou,M., Basil,M., Bradbury,P., Brinac,S., Bunesler,R., Davis,C., English,C., Franklin,T.L., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Haber,L., Harris,J., Hinson,S., Narayanaswamy,U., Newton,J., O'Brien,K., Patel,P., Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T. and Wilson,R.

TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
3 (bases 1 to 65020)
Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basil,M., Buetner,J., Butler,C., Card,P., desaiIboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Haber,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N., Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Walker,K. and Ward,T.

TITLE Direct Submission
JOURNAL Submitted (29-OCT-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
On Oct 29, 1998 this sequence contains the entire insert of clone PDJ254e13. PDJ254e13 comes from the RPCR-3 PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.85. In addition, attempts have been made to assure 98% of consensus base calls consist of either double-stranded coverage or 2 types of labelling chemistry on one strand.

FEATURES
source 1..65020
/organism="Homo sapiens"
/db_xref="taxon:9606"
repeat_region complement(5..86)
/rpt_family="THE1"
repeat_region 1040..1347
/rpt_family="Alu"
repeat_region 1543..1632
/rpt_family="THE1"
repeat_region complement(2645..2925)
/rpt_family="Alu"
repeat_region complement(3404..3692)
/rpt_family="Alu"
repeat_region 5568..5990
/rpt_family="Alu"
repeat_region complement(6925..7021)
/rpt_family="MIR"
repeat_region 9109..9172
/rpt_family="MLT1"
repeat_region complement(9452..9630)
/rpt_family="Alu"
repeat_region complement(10004..10079)
/rpt_family="MLT1"
repeat_region complement(10087..10280)
/rpt_family="MLT1"
repeat_region complement(10686..10959)
/rpt_family="Alu"
repeat_region complement(12119..12387)
/rpt_family="Alu"
repeat_region complement(15054..15256)
/rpt_family="Alu"
repeat_region complement(15762..16055)
/rpt_family="MER4.2"
repeat_region complement(16280..16476)
/rpt_family="L1"
repeat_region 17758..17984
/rpt_family="Alu"
repeat_region complement(18051..18258)

repeat_region /rpt_family="L1"
20160..20437
repeat_region /rpt_family="Alu"
22430..22738
repeat_region /rpt_family="Alu"
26073..26379
repeat_region /rpt_family="Alu"
complement(27905..28181)
repeat_region /rpt_family="Alu"
30140..30466
repeat_region /rpt_family="Alu"
32260..32543
repeat_region /rpt_family="Alu"
34385..34674
repeat_region /rpt_family="Alu"
complement(38169..38537)
repeat_region /rpt_family="THE1"
39230..39318
repeat_region /rpt_family="MIR"
40108..40247
repeat_region /rpt_family="MIR"
complement(40311..40580)
repeat_region /rpt_family="Alu"
40618..40714
repeat_region /rpt_family="L1"
46149..46227
repeat_region /rpt_family="MLT1"
47776..48090
repeat_region /rpt_family="Alu"
complement(51040..51183)
repeat_region /rpt_family="MERS"
complement(52705..52989)
repeat_region /rpt_family="Alu"
complement(53823..53944)
repeat_region /rpt_family="MIR"
54921..55201
repeat_region /rpt_family="Alu"
complement(55486..55783)
repeat_region /rpt_family="Alu"
60173..60468
repeat_region /rpt_family="Alu"
BASE COUNT 20406 a 11757 c 12727 g 20130 t
ORIGIN

Query Match 74.8%; Score 17.2; DB 11; Length 65020;
Best Local Similarity 86.4%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ATTACACACTGGCATGCCGAT 23
DB 61843 ATTACACACAGCATGCACAT 61864

RESULT 9
MMANNREC 5085 bp mRNA ROD 25-SEP-1996
DEFINITION M.musculus mRNA for macrophage mannose receptor.
ACCESSION 211974
NID 652997
VERSION 211974.1 GI:52997
KEYWORDS macrophage mannose receptor; mannose receptor.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sclerognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4467)
AUTHORS Harris,N., Super,M., Rits,M., Chang,G. and Ezekowitz,R.A.
TITLE Characterization of the murine macrophage mannose receptor: demonstration that the downregulation of receptor expression mediated by interferon-gamma occurs at the level of transcription

JOURNAL Blood 80 (9), 2363-2373 (1992)
MEDLINE 93043353

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 39838)
Lamerdin,J.E., McCready,P.M., Richardson,P., Burkhardt-Schultz,K., Gordon,L., Stillwagen,S., Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Sakladakis,G., Dangnanan,L., Ertler,A., Christensen,M., Georgescu,A., Avila,J., Liu,S., Altix,C., Andeise,T., Scott,D., Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Krommiller,B., Arellano,A., Sanders,C., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
Sequence analysis of a 6.4 Mb region in 19p13.3 between CDC34 and VAV1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 39838)
AUTHORS Lamerdin,J.E.
JOURNAL Direct Submission
Submitted (09-JUN-1999) Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
Map and sequence oriented from P telomere to centromere. Cosmid R26610 overlaps cosmid F25540 (AC006277) to the left from bases 1 to 2,676 of this accession, and overlaps cosmid R31236 from bases 30,971 to 39,838. Additional map and sequence information may be obtained at: <http://www-bio.llnl.gov/dbp/genome/genome.html>.

FEATURES
source
1..39838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R26610"
/chromosome="19"
/map="19p13.3 between CDC34 and VAV1"
/cell_line="5H2-B"
/clone_id="LL19NC03 R chromosome 19-specific cosmid library"
/note="Cosmid library constructed at LLNL from flow-sorted chromosomes from human-hamster hybrid 5H2-B, which carries chromosome 19 as its only human chromosome."
1..139
/note="BLASTN similarity to 257824 (98..256); match: 0.98, score: 5.9e-55; database searched: nt; H.sapiens CPG island DNA genomic MseI fragment, clone 198b12, reverse read cp9198b12.rta"
complement(264..476)
/note="BLASTN similarity to 257823 (1..213); match: 1, score: 1.3e-79; database searched: nt; H.sapiens CPG island DNA genomic MseI fragment, clone 198b12, forward read cp9198b12.frla"
join(<338..421,5631..5672,10312..10374,10466..10687,12641..12796,13018..13268,15326..15467,16915..17065,17502..17578,18383..18487)
/note="Hypothetical groucho-related gene product; Hypothetical protein exhibits 52% identity to Mus musculus groucho-related gene product (AF145957) over 298 residues, and 36% identity to Tlcz_HUMAN (Q04725) over 292 residues. CDS constructed from combination of Xgral1 predictions and EST alignments"
/codon_start=1
/evidence=not_experimental
/product="R26610.1"
/protein_id="AA038076.1"
/db_xref="PID:g5030439"
/db_xref="GI:5030439"
/translation="AVFTGLDEQKFGKCIANGWDHGDPLKIQDVAEHHKQIG NVLIIVESCQLOGFQSEEVPAESPFGTPOQKCTEDLSESSDITMAIRSSMLR PLGDNQPEQLOFMDKPEWFDLTELQWIFRGMILHGEYLATLASSPFRHF TCGRGKIKVSLTNGQVADRPESHLPICPGATLPICLLSNRSKSLITGYNLASIS VMDIAPSLHVKEDLPACALNCALNDANLNDANLAFASFTYQVWILKQSVVDK GYDPGVKSIYVKGNTWTGPDACLCRWQDRTINKPLEYQFKSQINSLSPQDWDVL LGANAGQOMLOSTSGQRHMYGCDISVILSVKFSPIQOMWASGMDDFLGYSPAGT KVEVEPMSPVTCDVSSNNRLVVTGSGEHASVYQITY"
1260..1858
/note="BLASTN similarity to A049067.1 (1..599); match:

0.98, score: 2.5e-237; database searched: month.nt; RPT-11-268021.T0 RPT-11 Homo sapiens genomic clone RPT-11-268021, genomic survey sequence"
1951..2245
repeat_region
/rpt_family="AluLo"
complement(2442..2734)
/rpt_family="AluSg"
complement(2753..3027)
/rpt_family="AluSg"
complement(3030..3050)
/rpt_family="AT_rich"
3142..3463
repeat_region
/rpt_family="AluSg"
complement(3464..3496)
/rpt_family="WAD1"
3937..4240
repeat_region
/rpt_family="AluSx"
4252..4553
/rpt_family="AluLo"
4557..4849
repeat_region
/rpt_family="AluSg"
5173..5462
/rpt_family="AluSx"
complement(5524..5581)
/rpt_family="MIR"
5739..6038
repeat_region
/rpt_family="AluSx"
complement(6041..6181)
/rpt_family="LINE2"
complement(6684..7318)
/rpt_family="LINE2"
complement(7953..8250)
/rpt_family="AluSx"
8452..8746
repeat_region
/rpt_family="AluSx"
complement(8884..9201)
/rpt_family="AluY"
9277..9574
repeat_region
/rpt_family="AluSx"
9614..9905
/rpt_family="AluY"
9915..10206
repeat_region
/rpt_family="AluSx"
11798..12087
/rpt_family="AluY"
13479..13595
/rpt_family="MIR"
13600..13696
/rpt_family="AluY/PIAM"
complement(13708..13803)
/rpt_family="AT_rich"
13804..14104
/rpt_family="AluY"
14105..14194
/rpt_family="(TA)n"
14204..14486
repeat_region
/rpt_family="AluSx"
14490..14564
/rpt_family="(CA)n"
14565..14855
repeat_region
/rpt_family="AluSx"
complement(14856..14959)
/rpt_family="(TA)n"
15537..15673
repeat_region
/rpt_family="AluSg/x"
15675..15964
/rpt_family="AluSp"
15972..16276
/rpt_family="AluYb"
16390..16701
repeat_region
/rpt_family="AluSx"
complement(16737..16812)
/rpt_family="MIR"

CDS

NGVDFIDISTHOYPTDIAFSTSSNMEAMAKARGELAEVKKALAEAYLPVYTEM
NNSPRDPYHIDPYDAATVKTITDIDIGCYATWTFDIFEEBGQSILPFGHF
GLTNINGIKPSPRAQIOLDKLNGERIEFEDEKSPITDIAVONESEITLVSNNV
PLSPIDPENIKVVLKGIENCREVEVEIDEDIANPNKRVLEMSPAYLNREQIEELIK
ASLKKREKVSQWGVNNNEITFDLSVLPHSVAVATIKNG"
5439...>6067
/note="orf 6"
/codon_start=1
/transl_table=11
/protein_id="AA23064.1"
/db_xref="PID:g144301"
/db_xref="GI:144301"
/translation="MGIRTKRKPNTETVLRKAKISTLLIMTIVLELTKMPSYKK
AYAVVLANQIVGVKDKTEKONLITOKKVEERHNDSTFLOSLOLKSIEGQIRE
TRDELANTIEGVKLVKRIAFNSKPYFENPQTPNNILNKLKRVYNDKASQA
KLEKVEIKRVVSPAIKVADEATLTKMFGDOVLEYVYKSGDTLMDP"

BASE COUNT 2230 a 787 c 1243 g 1807 t
ORIGIN 1 bp upstream of BamHI site.

Query Match 73.0%; Score 16.8; DB 1; Length 6067;
Best Local Similarity 90.0%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AATTCACACACTGGCATTGCC 20
|||||
Db 2561 AATTACACACTGGCATTCC 2542

RESULT 13
LOCUS AF005383/c DNA BCT 27-NOV-1997
DEFINITION Caldicellulosiruptor saccharolyticus putative transport protein
(XynG), putative transport protein (XynH), xylanase (XynF),
xylanase (XynE), xylanase (XynD), xylanase (XynA), acetylxylosidase
(XynC) and xylanase (XynB) genes, complete cds.

ACCESSION AF005383
NID AF005383
VERSION AF005383.1 GI:2645411
KEYWORDS
SOURCE
ORGANISM
Caldicellulosiruptor saccharolyticus.
Caldicellulosiruptor saccharolyticus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Thermococcales; Thermococcales; Thermococcales; Thermococcales;
1 (bases 1 to 18483)
Luthi, E., Love, D.R., McNulty, J., Wallace, C., Caughey, P.A., Saul, D.
and Bergquist, P.L.
Cloning, sequence analysis, and expression of genes encoding
xylan-degrading enzymes from the thermophile 'Caldicellulium
saccharolyticum'
Appl. Environ. Microbiol. 56 (4), 1017-1024 (1990)

JOURNAL MEDLINE 2 (bases 1 to 18483)
REFERENCE 7e'o.V.S. Jr., Gibbs, M.D., Saul, D.J. and Bergquist, P.L.
A cluster of genes involved in xylan degradation cloned from the
extreme thermophile Caldicellulosiruptor saccharolyticus
Appl. Environ. Microbiol. (1997) In press
JOURNAL 3 (bases 1 to 18483)
AUTHORS 7e'o.V.S. Jr., Gibbs, M.D., Saul, D.J. and Bergquist, P.L.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-1997) School of Biological Sciences, Macquarie
University, Balacakra Road, Sydney, NSW 2109, Australia
FEATURES
Source
1. 18483
/organism="Caldicellulosiruptor saccharolyticus"
/db_xref="taxon:44001"
<1. 937
/note="orf 1"
/codon_start=2
/transl_table=11
/protein_id="AA87367.1"
/db_xref="PID:g2645412"
/db_xref="GI:2645412"

gene

CDS

/translation="DPEGVILGSRNFILOKMIINFSQNKTKYVPDVFETKRSILFYPI
NVATISMARVPRKLYRSPATCTPDDDYVFLMLSPNVRGRVKAISPKLFTKNI
SISISIMLSDGEGTQINLQREKIEPEQIVGIVKRNHYLITLTPNGVITLSESLV
DIPISKMLSDGEGTQINLQREKIEPEQIVGIVKRNHYLITLTPNGVITLSESLV
TANNEBELTEICDYLSNRYKLMIDLKRVYELKPVLSKIKKKFLEHRLKEGLV
LVKFEGETEILIVYTSIDQVYNTVYKARFKYIE"
953...1873
/gene="XynG"
953...1873
/gene="XynG"
/note="similar to Agrobacterium radiobacter lacG (part of
the binding-protein-dependent transport system for
lactose): Swiss-Prot Accession Number P29823"
/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AA87368.1"
/db_xref="PID:g2645413"
/db_xref="GI:2645413"

gene

CDS

/translation="MIRNPFQIFSRKSLKOKREKAMRGHLYLPVYVIGLIVFIIP
VNSVYTFQKSLGDLSEFSGFVNIYAFKDPNRYNLSISITMLYEPYIV
FSIFVAVLLKDEKGRFARTIFFPVYIASGVYVILKENVAGSGSNVASTYIF
KAENLAVILNAGIPRMSYVDIVNKKIPDLTRSGVOILLIALLHNIPKSEFEYA
VIEGATEMEKEMKITEPMTSPYVVALIYSIDYFDYGNQVARNVDEANKRFEYS
TTIAIIVPMYIILLNRYIKRVYVLT"
1918...2937
/gene="XynH"
1918...2937
/gene="XynH"
/note="similar to Agrobacterium radiobacter lacG (part of
the binding-protein-dependent transport system for
lactose) encoded by the sequence presented in Genbank
Accession Number X66596"

CDS

/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AA87369.1"
/db_xref="PID:g2645414"
/db_xref="GI:2645414"
/translation="MRNVKSLMSYIGKRLNVEVTSVAKRRLVYLVKRVYAVY
CMYVILYPIFLIFLNSLRKEDLPDSVWIRKHTFSFEYANVLSYPAIKNTL
LMILPVIYIOTFISMWVGAFARERARLITLITLITFVLOIIVLYVYKPYF
DFYIGKLGFTIGKPLVYVLTTPPFTILNFGKIRSLIYTFPPOFRMPPEL
EBAKIDGCPSESTFIRINVPNAVATITVWLSIVMHWNDIVYSAFENLPISTM
LTVLNRMMSIOWAYNFSSEDIYLMGSLVLAQCLVPLIYIYIIGORYFESIER
TGIVG"
3016...4800
/note="orf 2"
/codon_start=1
/transl_table=11
/protein_id="AA87370.1"
/db_xref="PID:g2645415"
/db_xref="GI:2645415"
/translation="MRKSRKFKLVSTVLVCLITSVAVVISMVTOKASSTLGDIS
EUPGSRKESLKTDLFNRKAVTAIEDYOKKRGKRVIVSDNNONKIKIAMAAGD
PIDVITGSGVTFPSHFTRGVLPDLKIVYDLAKAPINIRADIVAFKNGHYIISQGS
NVPWLVYNNKDLMEGIDEEMPLALYKRGKNWOTPALAKRLADVTKDKIDRY
GVNFMATATVLYNGVAFVSDIKTGKGLIDFNAAQPALNFKRKAKESMLARDMDI
TVSGLRKQVYVAVOYKREDOKREVEDLEAPLQPDNKGKLVYFPADQYGLK
GSRNPVGAQKFTNLLESYOKNNDVYNAKRPVIVDVPVYKLAOKSPYPLGAGLGM
PHNDITGRVDSVSVALSSLRPQIEKNIKNDQOTREVEYVYKOFEPFTINFDGKXS
FKLIDTSKRTVRLSVSGKDAIKGSLKVTWDSKQKELIYVTDPAKVIYGMHDK
ISEFVYKAVTAGKTVSCITLISERPSSKSGKSNSTIKIDQOTVYKVGNTINIPD
NSOFMSLRIGITFGCVFIDNIRVEIE"
4881...8924
/gene="XynF"
4881...8924
/gene="XynF"
/function="has both arabinofuranosidase and exoxylanase
activities"
/codon_start=1
/transl_table=11
/product="xylanase"

```

/gene="xynA"
/function="endoxylanase activity"
/codon_start=1
/transl_table=11
/product="xylanase"
/protein_id="AAB87374.1"
/db_xref="GI:2645419"
/db_xref="PID:g2645419"
/translation="MANKLRLMKIKLVLPALIVTCLLSAFVNFQINKAYSSSSNOOG
SKLLQNTTSSNDPILSDIPDDIIRVNDGNDYRNTMTLPGVPMSTEDVNMRI
GYVYDLNDNDHNLKNGNMTYKGGKMGVGLGALDGTGDFYSSNNTI
TGPWEVEENETLHDSPLLFDNGKAYIIYGGSEIRKELMDIKRINKKNTYIK
SKIDCEESHAUKTKNGYITTTIRREKGIIECVYRDKIDGPEGVYSDTLTG
KINGVAQGLIDTPDGKMTAMLQDDAVGRVPLVPMWEGMPVEGDEKAPLKE
VRQGSSEFTEVYKSDYQKAEASNPREFVRKRNALIKDELVINNPPGTLGMK
GREGAKLEVEYSEGNKIINHYNKTGRSGIQNLGRVKKRKAFFVKKVGGD
KEFLIAKITSNGKESYOEILKGFANKNMTISGFTISODIETVYLITQIPYEP
DKNDLVDEYEVYSKYEYLATOKSEEAAPGNSVLTLMNNHNDNKSLLBERG
ELRKTQNVSDIOQARNLTORTLGRKSGNIMDYRMKGGDAGLAAPKEGFI
GVRKQKREYIVAVEKGEVEYTKLQSVYVLIKDEFDIDAVETYSNGINWIKFG
SOLJMKRTYIPHEMGYHFALENFATQGTGYVIDEFKRELDYFEALVADRKINAVL
IELRNDIDKNELVLSNLSPKGRKATQINILKFPRLDYEALVADRKINAVL
VEKGSODALINIKNIDTEVLDGSRNIDLVNIIKLRKREKLESTEDETRISLNEY
EKEINYSRESAVTKYFVAPKTPVCKIPNANPLIAKREGADPAVLAKNVTYILT
NDILEYDDNGNKNDYTKINKITITISSDDLNVPLVDEIGRPLSRSTPYEGVYVL
STAYKTKNGKDFLIFANNASGIGVLSDTVPGEDEIGRPLSRSTPYEGVYVL
FDPALVDDGKRAYIYFGGVQGDAMPNTARVQAGDMLSVGSAVTIPAPMFE
DSINKIGNTYIYSCITNFYSGARPQSGPAGVIAYMTSKSPGMEYKGVILKPN
FGVGGNNHGHLEFENGKMYIAYHQTOLAKDGVAKGYRSPHITOVEIENGKIREYA
DVRGIAOVYFPDPYKMYEAFETAMCAGISTKRNANSNCLTGIDSGDMLSTYDEG
NYGRKQTEAMVSNINGKIYEIKIDSYDCKIIGVYVQDSSTHMYVEYVNDKVT
GYHDYFVFKGSEKRLFDMDCKFKF"
/gene="xynB"
/gene="xynC"
/function="exoxylanase activity only"
/codon_start=1
/transl_table=11
/product="xylanase"
/protein_id="AAB87372.1"
/db_xref="PID:g2645417"
/db_xref="GI:2645417"
/translation="MKRTFKSVTLTAAVAVLVGCVLKAFPVQONAKQSSSTVTI
NEGKDTLTFYVGNKATIDQSSALEGRKIRVNRKNSIMSLADVDVYLRKRTW
VISSYIKHYGKPIAFSTIALYDNGKLYVGLGKIVPDSKEMIAVWKRVLKNS
NLIIAHPYDKTAVNDVSIQVNTVEETLSQAIYIKDFENNTNMRRGSGVILK
DSKFPHEGNEISYSGRAVFMHGAKPIIKYVVPGRGFSIMVHTSMDLRFSILY
ORKMADEAGRYRDMYTSKVEAGDGWELISJELIELEISDESPDLSFYV
DPTISDTYKIDQPNYSLPSLEKRYKNDKRYVAGYGLINSTOKPITKRPNSTP
GMEKPESLRGPDKYDFTIADAFVEFAIKNNISIRGHTLVHNOTPMEFFDSNGE
LKKDLKLRLKRIHYVGRYKGIYAMDVNEAIDEDPDYRRSNMVKTCGPYIE
KAFIWAHEADPOAKLEIYNDYNTVEPOKRMFIYMLKNNKSKPIHIGYLOCHINVS
PSVEIEETIRLESTIPGLEIOTELDNGSYFGMSSVYVVERELMLROAKRYELF
NFRYKRIKISVTFMGLKDDYSMLRGVYNNKDPFLDEYIDGAPAFMALIDYSVIP
ONINLPAPPAIRKMLTKK"
/gene="xynD"
/gene="xynE"
/function="xylosidase activity"
/codon_start=1
/transl_table=11
/product="xylanase"
/protein_id="AAB87373.1"
/db_xref="PID:g2645418"
/db_xref="GI:2645418"
/translation="MBEGIVTNKIKIEKKQIIFPDKKKFCVSGRIGLAKQETIDA
LIVKRIHDKFIRAHGLHDDVGITREDIDGNEVSFNFTYIDRTIDSLFELDIRP
FVEIGFMPKSLASGTQVYVWRGNANTPPADYKWRMLIKSVKHEIDRGEVEVOMP
FEIINEPNTLVNFKMDANQAEYFKLEYAKAKAEVNEIKVGPALICGSDYMDIDL
HCYKKNYVDELTRHAATYGRKPPQATYVHVVDYVPIEMLKEFSYRSEKVAEPPD
LPHITEENSSYHPICTHDPDPNAYLARTVSEAGDGYPSVYVVERELMLROAKRYELF
AIFHGFGIVANNNIPKPYFHFTFPNMGAKILYRDEHILITTEBNGIATVANNRY
MKKEATDKRIEILPVDINDVFIKOKLIDEEBNGMWRKTIOMG"
/gene="xynA"
/gene="xynB"
/gene="xynC"

```

```

/gene="xynA"
/function="endoxylanase activity"
/codon_start=1
/transl_table=11
/product="xylanase"
/protein_id="AAB87374.1"
/db_xref="GI:2645419"
/db_xref="PID:g2645419"
/translation="MRLIVCEENLEMLNLSAKTYKDYFKIGAAVYANDLEGVHNDIL
LKHFPSLPEPMKKEFNHPEORYNEFEVARIKEFAIKNDMLGHTFVWHNOTPGM
VFLDNGEASKEIVIERLRHRIKTCRYKQVYVAMDVNEAEDKTKLLRESNMR
KITGDYIKIAFEIAREVAGDAKLEFYNDNNMPPKLEKTKVLEKELERPTPIDGIS
Query Match
Best local similarity 73.0%; Score 16.8; DB 2; Length 18483;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 14977 AATTTCACACTGGCATGCC 14958
Query 1 AATTTCACACTGGCATGCC 20
|||||
Db 14977 AATTTCACACTGGCATGCC 14958
RESULT 14
LOCUS 184210 2062 bp DNA PAT 20-MAR-1998
DEFINITION Sequence 8 from patent US 5695766.
ACCESSION 184210
MID 93021730
VERSION 184210.1 GI:3021730
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2062)
AUTHORS Paul, P.S., Halbur, P.G., Meng, X., Lyoo, Y.S. and Lum, M. Anne.
TITLE Highly virulent porcine reproductive and respiratory syndrome
viruses which produce lesions in pigs and vaccines that protect
pigs against said syndrome
Patent: US 5695766-A 8 09-DEC-1997;
FEATURES
source location/Qualifiers
1..2062
BASE COUNT 480 a 491 c 530 g 561 t
ORIGIN
Query Match
Best local similarity 73.0%; Score 16.8; DB 5; Length 2062;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Query 1 AATTTCACACTGGCATGCC 20
|||||
Db 1935 AATTTCACACTGGCATGCC 1916
RESULT 15
LOCUS 184217 164 bp DNA PAT 20-MAR-1998
DEFINITION Sequence 19 from patent US 5695766.
ACCESSION 184217
MID 93021737
VERSION 184217.1 GI:3021737
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 164)
AUTHORS Paul, P.S., Halbur, P.G., Meng, X., Lyoo, Y.S. and Lum, M. Anne.
TITLE Highly virulent porcine reproductive and respiratory syndrome
viruses which produce lesions in pigs and vaccines that protect
pigs against said syndrome
Patent: US 5695766-A 19 09-DEC-1997;
FEATURES
location/Qualifiers

```


source 1.164
BASE COUNT 50 a /organism="unknown"
ORIGIN 28 c 45 g 41 t

Query Match 73.0% Score 16.8; DB 5; Length 164;
Best Local Similarity 90.0% Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AATTCACACACTGGCATGCC 20
|||||
Db 37 AATTCACACACTGGCATGCC 18

Search completed: September 13, 1999, 15:55:19
Run time: 4547 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:29 ; Search time 255.05 Seconds

(without alignments)
22.562 Million cell updates/sec

Title: US-09-325-095-15

Perfect score: 23

Sequence: 1 AATTGACACACTGGCATGCCGAT 23

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	23	100.0	3207	1	051155	p110 CDNA. Recombi
2	23	100.0	23	1	039014	RACE PCR primer 22
3	23	100.0	3498	1	057012	Pitins 3-kinase 11
4	23	100.0	23	1	057020	Pitins 3-kinase pr
5	21.4	93.0	3412	1	051156	Human p110 CDNA. R
6	16.8	73.0	1938	1	063591	ISF-12 3' terminal
7	16.8	73.0	164	1	063598	ISF-12 non-transla
8	16.8	73.0	164	1	114394	PRSV VR 2385 3'-n
9	16.8	72.2	2062	1	114389	PRSV VR 2385 3' t
10	16.6	72.2	5559	1	055260	Restriction fragme
11	16.6	72.2	5559	1	T90945	E11 genomic DNA fr
12	16.6	72.2	5559	1	T96101	Pseudomonas fluore
13	16.6	72.2	5559	1	T89212	5.6 kb E11 Pseudom
14	16.6	72.2	5559	1	V06486	3.6 kb EcoRI-Hind
15	16.6	72.2	10763	1	V39837	Pseudomonas fluore
16	15.6	67.8	7430	1	V74571	Staphylococcus aur
17	15.4	67.0	2680	1	V64579	Myc-binding zinc-f
18	15.4	67.0	150	1	X12033	Human biallelic po
19	15.4	67.0	203	1	X41456	Human secreted pro
20	15.2	66.1	2375	1	032001	Encodes elmeria an
21	15.2	66.1	4374	1	Q72914	Murine receptor ty
22	15.2	66.1	3358	1	T16238	VR-2332 Viral sequ
23	15.2	66.1	298	1	T24074	Human gene sigma
24	15.2	66.1	2375	1	V13334	Elmeria acerulina
25	15.2	66.1	451	1	V74934	Staphylococcus aur
26	15.2	65.1	2330	1	X34729	Ethylene receptor
27	15	65.2	1247	1	N50520	Sequence of brain
28	15	65.2	3207	1	051155	p110 CDNA. Recombi
29	15	65.2	3498	1	057012	Pitins 3-kinase 11
30	15	65.2	2058	1	062597	Pitins 3-kinase 11
31	15	65.2	3913	1	T05573	Human cytoplasmic
32	15	65.2	1266	1	T67467	MEKK4 CDNA. New sl
33	15	65.2	1278	1	T68187	H. pylori transmem
34	15	65.2	690	1	V24807	H. pylori ORF hp3p
35	15	65.2	528	1	V24941	H. pylori cellular
36	15	65.2	2393	1	V13443	Probe HP-60 used t
37	15	65.2	3913	1	V22679	CDNA encoding a mi
38	15	65.2	5399	1	V47932	DNA encoding prote
39	15	65.2	3913	1	V45617	MEKK4 protein codi
40	15	65.2	5233	1	X13150	Enterococcus faeca
41	15	65.2	50000	1	X23517	Human kidney amino
42	14.8	64.3	2297	1	V07896	Cobra venom moca
43	14.8	64.3	2288	1	V07898	Cobra venom moca

ALIGNMENTS

44	14.8	64.3	2309	1	V07899	Cobra venom moca
45	14.8	64.3	1820	1	V07900	Cobra venom moca

RESULT	1
051155/c	
ID	051155 standard; CDNA; 3207 BP.
AC	051155:
DT	12-APR-1994 (first entry)
DE	p110 CDNA.
KW	phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW	ss.
FT	Key
FT	Location/Qualifiers
FT	1..3207
FT	/*tag= a
FT	/note= "PI3- kinase p110"
PN	MO9321328-A.
PD	28-OCT-1993.
PE	13-APR-1993; G00761.
PR	13-APR-1992; GB-008135.
PA	(LUDW-) LUDWIG INST CANCER RES.
PI	Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI	Parkerp, Volinia S, Waterfield MD;
DR	WPI: 93-351738/44.
DR	P-RSD: R43341.
PT	Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT	activity, useful for controlling cell proliferation
PS	Claim 7; Fig 9; 146pp; English.
CC	Southern blot analysis was performed using a bovine CDNA probe conty.
CC	a fragment of a PI3-kinase-encoding sequence and human CDNA isolated
CC	from a CDNA library constructed from mRNA isolated from the human
CC	cell line K562. Positive clones were sequenced to give the human
CC	PI3 kinase p110 sequence shown. This sequence has 95 percent
CC	homology with the bovine sequence. The domain encoding residues 19-
CC	100 of human p110 is sufficient to encode the kinase which will
CC	associate with the p85 kinase subunit. The gene may be used to
CC	provide a protein with PI3 kinase activity, and is useful for
CC	screening for (ant)agonists of PI3 kinase activity which could be
CC	useful for stimulation or inhibition of cell proliferation and hence
CC	prophylaxis or therapy. Platelet or neutrophil activity or blood
CC	glucose levels can be controlled using the kinase.
CC	See also 051156, 059012-23 and 057522-3.
SQ	Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match	100.0%; Score 23; DB 1; Length 3207;
Best Local Similarity	100.0%; Pred. No. 0.014;
Matches	23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 AATTGACACACTGGCATGCCGAT 23
DB	383 AATTGACACACTGGCATGCCGAT 361

RESULT	2
059014	
ID	059014 standard; DNA; 23 BP.
AC	059014:
DT	12-APR-1994 (first entry)
DE	RACE PCR primer 2224.
KW	phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW	ss; amplification; p110.
OS	Synthetic.
PN	MO9321328-A.
PD	28-OCT-1993.
PF	13-APR-1993; G00761.

PR 13-APR-1993: GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1; Page 38; 146pp; English.
 CC An SGBAF-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with pSV3neo. Total RNA was
 CC isolated from the SGBAF-1 and a cDNA library constructed. RACE PCR
 CC was performed by synthesizing first strand cDNA from random hexamers
 CC on SGBAF-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-
 CC dt as primers. Products were fractionated using an agarose gel. The
 CC cDNA was isolated from the gel and subjected to PCR using oligo 2280 and
 CC adaptor as primers. A product of 350 bp was further sequenced.
 CC See also 051155-6, 059012-23 and 057522-3.
 SO Sequence 23 BP; 7 A; 7 C; 4 G; 5 T;

Query Match 100.0%; Score 23; DB 1; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.0065;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCCGAT 23
 Db 1 AATTCACACACTGGCATGCCGAT 23

RESULT 3
 057013/c
 ID 057012 standard; cDNA to mRNA; 3498 BP.

AC 057012;
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase 110 kD catalytic subunit cDNA.
 KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schistosaccharomyces pombe; nmt promoter; thiamine;
 KW ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Bos taurus.
 FH Key
 FT cds 1.3207
 FT /tag- a
 FT /product- p110

PN WO9403609-A.
 PD 17-FEB-1994.
 PR 05-AUG-1993; G01651.
 PR 05-AUG-1993; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 DR P-PsDB; R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure: Fig 1; 71pp; English.
 CC This sequence encodes the 110 kD catalytic subunit of the phosphatidylinositol (Ptdins) 3-kinase. This sequence was transformed into Schistosaccharomyces pombe cells under the regulatory control of the nmt promoter in an embodiment of the invention. In the presence of thiamine the promoter is inactive and the cells carrying the Ptdins catalytic subunit plasmid grow as the parental strain. In the absence of thiamine the nmt promoter functions and the Ptdins 3-kinase catalytic subunit is induced. Ptdins activity is substantially increased under these conditions. Cells containing constructs such as this, are useful in assays for detecting compounds involved in cell growth regulation. It is also used as the basis for detecting compounds for treating cancers and the formation of blood vessel plaques.

Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match

100.0%; Score 23; DB 1; Length 3498;

Best Local Similarity 100.0%; Pred. No. 0.014;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCCGAT 23
 Db 383 AATTCACACACTGGCATGCCGAT 361

RESULT 4

ID 057020 standard; DNA; 23 BP.
 AC 057020;
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase primer oligo 2224.
 KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schistosaccharomyces pombe; nmt promoter; thiamine;
 KW ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Synthetic.
 PN WO9403609-A.
 PD 17-FEB-1994.
 PR 05-AUG-1993; G01651.
 PR 05-AUG-1993; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Example 1; Page 15; 71pp; English.

CC The sequences given in 057020-23 are primers which were used in the
 CC isolation of the phosphatidylinositol (Ptdins) 3-kinase cDNA. The
 CC amplified sequence was placed under the regulatory control of the nmt
 CC promoter. These sequences were transformed into Schistosaccharomyces
 CC pombe cells in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the Ptdins
 CC plasmids grow as the parental strain. In the absence of thiamine the
 CC nmt promoter functions and the Ptdins is induced. Ptdins activity is
 CC substantially increased under these conditions. Cells containing
 CC constructs such as this, are useful in assays for detecting compounds
 CC involved in cell growth regulation. It is also used as the basis for
 CC detecting compounds for treating cancers and the formation of blood
 CC vessel plaques.
 SO Sequence 23 BP; 7 A; 7 C; 4 G; 5 T;

Query Match 100.0%; Score 23; DB 1; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.0065;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCCGAT 23
 Db 1 AATTCACACACTGGCATGCCGAT 23

RESULT 5

ID 051156/c
 AC 051156;
 DT 12-APR-1994 (first entry)
 DE Human p110 cDNA.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neurorphil activity; 3-phosphorylated phosphoinositides;
 OS Human.

FH Key
 FT cds 1.3207
 FT /tag- a
 FT /note- "P13- kinase p110"

PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.

PR 13-APR-1992: GB-008135.
PA (LUDM-) LUDMIG INST CANCER RES.
PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI Parker PJ, Volinia S, Waterfield MD;
DR WPI: 93-351738/44.
DR P-PSDB: R43342.
PT Recombinant polypeptide(s) - with phosphonostide-3 kinase
PT activity, useful for controlling cell proliferation
PS Claim 7; Fig 16: 146pp; English.
CC Southern blot analysis was performed using a bovine cDNA probe contg.
CC a fragment of a p13-kinase-encoding sequence and human cDNA isolated
CC from a cDNA library constructed from mRNA isolated from the human
CC cell line KG1a. Positive clones were sequenced to give the human
CC p13 kinase p10 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p10 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with p13 kinase activity, and is useful for
CC screening for (ant)agonists of p13 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also Q51155 and Q57522-3.
CC Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;
CC
CC
Query Match 93.0%; Score 21.4; DB 1; Length 3412;
Best Local Similarity 95.7%; Pred. No. 0.088;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 AATTGACACACTGGCATGCCGAT 23
DB 383 AATTGACACACTGGCATGCCGAT 361
RESULT 6
O63591/C
AC 063591 standard; cDNA: 1938 BP.
DT 12-DEC-1994 (first entry)
DE ISU-12 3' terminal sequence.
KW Primer: polymerase chain reaction; PCR; amplify; probe: Iowa strain;
KW infectious agent; porcine respiratory and reproductive syndrome; PRD;
KW ISU-12; vaccine; porcine respiratory and reproductive disease; PRD;
KW antibody; assay; ss.
OS Porcine respiratory and reproductive disease virus, ISU-12.
FS Key Location/Qualifiers
FT cds 239..901
FT /*tag- a
FT /label- ORF-5
FT cds 889..1403
FT /*tag- b
FT /label- ORF-6
FT cds 1403..1771
FT /*tag- c
FT /label- ORF-7
FT
PM EP-595436-A.
PD 04-MAY-1994.
PF 29-OCT-1993: 203042.
PR 30-OCT-1993: US-969071.
PR 05-OCT-1993: US-131625.
PA (SOLV) SOLVAT ANIMAL HEALTH INC.
PA (IOWA) UNIV IOWA STATE RES FOUND INC.
PI Halbur PG, Lum MA, Lyoo XS, Meng X, Paul PS;
DR WPI: 94-146025/18.
DR P-PSDB: R53462.
PT New porcine respiratory and reproductive disease virus - used to
PT prepare vaccines and antibodies for diagnosis, treatment and
PT prophylaxis of virus infection.
PS Example 4; Page 45-47; 98pp; English.
CC This sequence represents the 3' terminal region of the infectious
CC agent associated with the Iowa strain of porcine respiratory and
CC reproductive syndrome; termed ISU-12. Three large open reading

CC frames (ORFs) have been found in this sequence, ORF-5, ORF-6 and
CC ORF-7 (see also Q63592-94). These ORFs all have the capacity of
CC encoding more than 100 amino acids. ORF-5 and ORF-6 overlap each
CC other by 10 bp and ORF-6 and ORF-7 overlap each other by 5 bp. This
CC isolated ISU-12 sequence may be used to infect cells and from these,
CC the vaccine of the invention can be produced. This vaccine may be
CC used for protecting pigs against a porcine respiratory and reproductive
CC disease (PRD). Antibodies to the vaccine may also be used in treating
CC PRD and for assaying for the virus.
SQ Sequence 1938 BP; 453 A; 458 C; 497 G; 530 T;
Query Match 73.0%; Score 16.8; DB 1; Length 1938;
Best Local Similarity 90.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AATTGACACACTGGCATGCC 20
DB 1811 AATTGACACACTGGCATGCC 1792
RESULT 7
O63598/C
ID 063598 standard; cDNA: 164 BP.
AC 063598;
DT 12-DEC-1994 (first entry)
DE ISU-12 non-translated sequence.
KW Primer: polymerase chain reaction; PCR; amplify; probe: Iowa strain;
KW infectious agent; porcine respiratory and reproductive syndrome;
KW ISU-12; vaccine; porcine respiratory and reproductive disease; PRD;
KW antibody; assay; ss.
OS Porcine respiratory and reproductive disease virus, ISU-12.
FS EP-595436-A.
PD 04-MAY-1994.
PF 29-OCT-1993: 203042.
PR 30-OCT-1993: US-969071.
PR 05-OCT-1993: US-131625.
PA (SOLV) SOLVAT ANIMAL HEALTH INC.
PA (IOWA) UNIV IOWA STATE RES FOUND INC.
PI Halbur PG, Lum MA, Lyoo XS, Meng X, Paul PS;
DR WPI: 94-146025/18.
PT New porcine respiratory and reproductive disease virus - used to
PT prepare vaccines and antibodies for diagnosis, treatment and
PT prophylaxis of virus infection.
PS Example 4; Page 53; 98pp; English.
CC This sequence represents the non-translated sequence from the 3'
CC terminal region of the infectious agent associated with the Iowa
CC strain of porcine respiratory and reproductive syndrome, termed
CC ISU-12. This sequence ends the ISU-12 fragment which contains open
CC reading frames (ORFs) ORF-5, ORF-6 and ORF-7. These ORFs all have
CC the capacity of encoding more than 100 amino acids. ORF-5 and ORF-6
CC overlap each other by 10 bp and ORF-6 and ORF-7 overlap each other by
CC 5 bp. The isolated ISU-12 3' sequence may be used to infect cells and
CC from these infected cells the vaccine of the invention can be produced.
CC This vaccine may be used for protecting pigs against a porcine
CC respiratory and reproductive disease (PRD). Antibodies to the
CC vaccine may also be used in treating PRD and for assaying for the
CC virus.
SQ Sequence 164 BP; 50 A; 28 C; 45 G; 41 T;
Query Match 73.0%; Score 16.8; DB 1; Length 164;
Best Local Similarity 90.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AATTGACACACTGGCATGCC 20
DB 37 AATTGACACACTGGCATGCC 18
RESULT 8
T14394/C
ID 114394 standard; cDNA: 164 BP.

AC T14394; 05-AUG-1996 (first entry)
 DE PRRSV VR 2385 3'-nontranslational region.
 KM PRRSV; vaccine; antigen; ss.
 OS pig reproductive and respiratory syndrome virus Iowa (VR 2385).
 PN M09606619-A1.
 PD 07-MAR-1996.
 PE 01-SEP-1995: U10904.
 PR 01-SEP-1994: US-301435.
 PA (HALB/) HALBUR P.
 PA (LUMM/) LUM M A.
 PA (MENG/) MENG X.
 PA (MORO/) MOROZOV I.
 PA (PAUL/) PAUL P S.
 PI Halbur P, Lum MA, Meng X, Morozov I, Paul PS;
 DR WPI: 96-160132/16.
 PT New porcine reproductive and respiratory syndrome virus DNA - and
 PT proteins encoded by open reading frames of an Iowa strain of the
 PT virus; are used in vaccines against PRRSV in pigs
 PS Disclosure; Page 147: 228pp; English.
 CC The 3'-nontranslational sequence following the stop codon of
 CC ORF-7 (see also T14392) of porcine reproductive and respiratory
 CC syndrome virus (PRRSV) strain Iowa isolate ISU-12 (VT 2385) was
 CC identified in the genome of VR 2385 and was compared to the
 CC corresponding sequence (T14397) of PRRSV Ielystad, a European strain.
 CC A consensus region adjacent to the polyA tail was identified. 588
 CC homology was found between the 2 sequences.
 SC Sequence 164 BP; 50 A; 28 C; 45 G; 41 T;

Query Match 73.0%; Score 16.8; DB 1; Length 164;
 Best Local Similarity 90.0%; Pred. No. 11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCC 20
 ||||| ||||| |||||
 DB 37 AATTCACACACTGGCATGCC 18

RESULT 9
 T14389/C
 ID T14389 standard; cDNA; 2062 BP.
 AC T14389;
 DT 04-AUG-1996 (first entry)
 DE PRRSV VR 2385 3' terminal sequence.
 KM PRRSV; vaccine; antigen; ss.
 OS pig reproductive and respiratory syndrome virus Iowa strain ISU-12.
 FH Key location/Qualifiers
 FT cds 426..1028
 FT /*tag- a
 FT /label- ORF-5
 FT 1013..1037
 FT /*tag- b
 FT /label- ORF-6
 FT /product- viral membrane protein
 FT 1527..1898
 FT /*tag- c
 FT /label- ORF-7
 FT /product- capsid protein
 PN M09606619-A1.
 PD 07-MAR-1996.
 PE 01-SEP-1995: U10904.
 PR 01-SEP-1994: US-301435.
 PA (HALB/) HALBUR P.
 PA (LUMM/) LUM M A.
 PA (MENG/) MENG X.
 PA (MORO/) MOROZOV I.
 PA (PAUL/) PAUL P S.
 PI Halbur P, Lum MA, Meng X, Morozov I, Paul PS;
 DR WPI: 96-160132/16.
 DR P-PSDB: R94701;
 DR P-PSDB: R94702;
 DR P-PSDB: R94703;

PT New porcine reproductive and respiratory syndrome virus DNA - and
 PT proteins encoded by open reading frames of an Iowa strain of the
 PT virus; are used in vaccines against PRRSV in pigs
 PS Disclosure; Page 139-140: 228pp; English.
 CC The 3' terminal sequence (T14389) of a porcine reproductive and
 CC respiratory syndrome virus (PRRSV) Iowa strain isolate ISU-12 (VR
 CC 2385) cDNA clone was detd. following PCR amplification (see also
 CC T14377-83). The sequence includes ORFs 5 (see also T14390), 6
 CC (T14391) and 7 (T14391), whose products (R94701-03) can be used in
 CC the development of pig PRRSV vaccines.
 SC Sequence 2062 BP; 480 A; 491 C; 530 G; 561 T;

Query Match 73.0%; Score 16.8; DB 1; Length 2062;
 Best Local Similarity 90.0%; Pred. No. 16;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCC 20
 ||||| ||||| |||||
 DB 1935 AATTCACACACTGGCATGCC 1916

RESULT 10
 ID 055260
 AC 055260;
 DT 28-JUL-1994 (first entry)
 DE Restriction fragment comprising gene activating sequence.
 KM Gene activation; gafa; transcriptional activator; activator;
 KM sensor; regulation; induction; biocontrol agents; crop protection;
 KM ss.

OS Pseudomonas fluorescens.
 FH Key location/Qualifiers
 FT cds 210..1688
 FT /*tag- a
 FT /label- ORF 1.
 FT 1906..3633
 FT /*tag- b
 FT /label- ORF 2.
 FT 4616..4691
 FT /*tag- c
 FT /label- g1yw
 FT /note- tRNA coding sequence.
 FT cds 4731..5318
 FT /*tag- d
 FT /label- ORF 3.

PN M09401561-A.
 PD 20-JAN-1994.
 PE 02-JUL-1993: U06300.
 PR 02-JUL-1992: US-908284.
 PA (CIBA) CIBA GEIGY AG.
 PI Becker J, Gaffney TD, Hill DS, Howell CR, Lam ST;
 PI Lyon JM, Stein JI;
 DR WPI: 94-035067/04.
 PT Gene activating sequences, partlc. from Pseudomonas - used for
 PT inducing expression of genes which are latent or expressed at low
 PT levels in target organisms
 PS Disclosure; Page 71-74: 87pp; English.
 CC The gene activating sequence is contained within an 11 kilobase
 CC EcoRI fragment (E11) from P. fluorescens. DNA sequence analysis of
 CC E11 has identified five open reading frames (ORF), of which ORFs
 CC and ORF 2 share significant homology with numerous activator and
 CC sensor components respectively of bacterial two-component
 CC regulatory systems. The gene activating sequences can be used in
 CC microorganisms to induce the production of gene products and
 CC secondary metabolites. They are used particularly for production
 CC of bacterial biocontrol agents for use in inhibiting plant
 CC pathogens. This sequence is a fragment of E11. See also 055261.
 SC Sequence 5559 BP; 1194 A; 1743 C; 1637 G; 984 T;

Query Match 72.2%; Score 16.6; DB 1; Length 5559;
 Best Local Similarity 82.6%; Pred. No. 23;

PT US5686283-A. /product- ORF3 protein
 PN 11-NOV-1997.
 PR 02-JUN-1995; 460298.
 PR 08-AUG-1994; US-287442.
 PR 20-AUG-1990; US-570184.
 PR 02-JUL-1992; US-908284.
 PR 01-JUL-1993; US-087636.
 PR 08-JUN-1994; US-258261.
 PR 02-JUN-1995; US-460298.
 PA (NOVS) NOVARTIS FINANCE CORP.
 PI Gaffney TD, Lam ST;
 PI MPI: 97-558142/51.
 DR Pseudomonas fluorescens sigma factor rpos gene - useful for
 PT activating gene expression in Pseudomonads
 PS Example 3; Column 41-46; 35pp; English.
 CC This sequence represents a 5.6 kb EcoRI genomic fragment from Pseudomonas
 CC fluorescens. The ORF5 (gafa) coding sequence (see T99211) can be linked
 CC to rpos, which is a gene-activating element capable of inducing
 CC the expression of at least one gene that is latent or expressed at low
 CC levels in a Pseudomonad. The P. fluorescens rpos gene can be used to
 CC stimulate biocontrol factors and secondary metabolites in bacteria,
 CC especially Pseudomonas spp. The expression pattern of rpos can be altered
 CC by placing it behind bacterial regulatory elements. Altering the temporal
 CC expression of rpos could induce earlier expression of rpos-regulated
 CC genes, which would increase production of biocontrol factors such as
 CC pyrrolinidin, chitinase, cyanide and gelatinase and thus render the host
 CC more effective for biological control of fungal pathogens.
 CC Sequence 5559 BP; 1194 A; 1747 C; 1633 G; 984 T;
 SO

Query Match 72.2%; Score 16.6; DB 1; Length 5559;
 Best Local Similarity 82.6%; Pred. No. 23;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AATTCACACACTGGCATGCCGAT 23
 DB 3394 ATTTCAGACATCGCATGCCGAT 3416
 1 ||||| ||| ||||| |||||

RESULT 14
 V06486
 ID V06486 standard; DNA; 5559 BP.
 AC V06486;
 DT 01-MAY-1998 (first entry)
 DE 5.6 kb EcoRI-Hind III restriction fragment of E11.
 KW Pseudomonas; gafa; transcription; latent gene; antifungal; E11;
 KW gene activating element; enzyme; plant pathogen; restriction; ss.
 OS Pseudomonas fluorescens.
 FH Key
 FT misc_feature 210..1688
 FT Location/Qualifiers
 FT /*tag- a
 FT /note- "ORF 1, transcribed left to right"
 FT 1906..3633
 FT /*tag- b
 FT /note- "ORF 2, transcribed left to right"
 FT 4616..4691
 FT /*tag- c
 FT /note- "gly W, transcribed right to left"
 FT 4731..5318
 FT /*tag- d
 FT /note- "ORF 3, transcribed right to left"
 FT
 FT US5710031-A.
 PN 20-JAN-1998.
 PE 02-JUN-1995; 459174.
 PR 08-AUG-1994; US-287442.
 PR 20-AUG-1990; US-570184.
 PR 02-JUL-1992; US-908284.
 PR 01-JUL-1993; US-087636.
 PR 08-JUN-1994; US-258261.
 PR 02-JUN-1995; US-459174.
 PA (NOVS) NOVARTIS FINANCE CORP.
 PI Gaffney TD, Hill DS, Lam ST, Ligon JW, Stein JT;

DR MPI: 98-109819/10.
 PR New Pseudomonas gafa element - used to activate transcription of
 PT latent genes or those expressed at low level, useful for, e.g.
 PT inducing synthesis of antifungal agents
 PS Example 5; Columns 41-46; 35pp; English.
 CC This is the DNA sequence of a 5.6 kb EcoRI-Hind III restriction fragment
 CC of fragment 11. A 2.0kb XhoI subclone of fragment E11 contains the open
 CC reading frame (ORF) 5 which encodes for a Pseudomonas gafa sequence. This
 CC gafa sequence is a gene activating element that can induce expression of
 CC at least 1 gene in a transformed Pseudomonas where the gene is latent or
 CC expressed at low level. The gene activating element comprises a
 CC nucleotide sequence of gafa class of transcriptional regulators and can
 CC be introduced into a Pseudomonas host cell to induce the expression of a
 CC latent gene. This renders the Pseudomonas host effective against fungi
 CC pathogenic on plants. The transformants express pyrrolinidin, cyanide,
 CC chitinase and gelatinase which are active against Rhizoctonia solani,
 CC Helminthosporium graminiae and some species of Pythium and Fusarium.
 CC The gene activating element may also be used to isolate genes encoding
 CC anti-pathogenic substances (APS), active against fungi, bacteria,
 CC nematodes and viruses, and very generally to transform cells other than
 CC Pseudomonas for increasing production of, e.g. antibiotics for
 CC pharmaceutical use, vitamins, growth factors, hormones, butanol, lactic
 CC acid, polysaccharides and enzymes. Transformed cells of Pseudomonas are
 CC applied at 50-5000 (especially 200-500) g/hectare. Transformed
 CC Pseudomonas have a broad spectrum of activity against plant pathogens
 CC and can compete more aggressively in the rhizosphere.
 CC Sequence 5559 BP; 1194 A; 1747 C; 1633 G; 984 T;
 SO

Query Match 72.2%; Score 16.6; DB 1; Length 5559;
 Best Local Similarity 82.6%; Pred. No. 23;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AATTCACACACTGGCATGCCGAT 23
 DB 3394 ATTTCAGACATCGCATGCCGAT 3416
 1 ||||| ||| ||||| |||||

RESULT 15
 V39837
 ID V39837 standard; DNA; 10763 BP.
 AC V39837;
 DT 29-SEP-1998 (first entry)
 DE Pseudomonas fluorescens strain CGA267356 11 kb EcoRI fragment.
 KW Pseudomonas; genetic engineering; biocontrol; plant; pathogenic;
 KW Rhizoctonia; Pythium; antifungal; pyrrolinidin; crop protection; ss.
 OS Pseudomonas fluorescens.
 FH Key
 FT CDS 210..1688
 FT Location/Qualifiers
 FT /*tag- a
 FT /product- "methyltransferase"
 FT /note- "coding sequence for methyltransferase has
 FT homology to the cher and frp genes from E.coli
 FT and Myxococcus xanthus, respectively"
 FT 1906..3633
 FT /*tag- b
 FT /product- "sensor kinase"
 FT /note- "coding sequence for sensor kinase has homology
 FT to the rcsC, frzE and bys genes of E.coli,
 FT M.xanthus and Bordetella pertussis, respectively"
 FT complement (4616..4691)
 FT /*tag- c
 FT /product- "tRNA"
 FT /note- "(complementary DNA strand) homology to glyW
 FT from E.coli"
 FT complement (4731..5318)
 FT /*tag- d
 FT /product- "CDP-diacylglycerol-glycerol-3-phosphate-3-
 FT phosphatidyltransferase"
 FT /note- "coding sequence for CDP-diacylglycerol-glycerol-
 FT 3-phosphate-3-phosphatidyltransferase has
 FT homology to pgsA"
 FT complement (5574..7397)
 FT CDS

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:55 ; Search time 1694.61 Seconds

(without alignments)
26.772 Million cell updates/sec

Title: US-09-325-095-15

Perfect score: 23
Sequence: 1 AATTCACACACTGGCATGCCGAT 23

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database:

EST:*

1: em_est1:*

2: em_est2:*

3: em_est3:*

4: em_est4:*

5: em_est5:*

6: em_est6:*

7: em_est7:*

8: em_est8:*

9: em_est9:*

10: em_est10:*

11: em_est11:*

12: em_est12:*

13: em_est13:*

14: em_est14:*

15: em_est15:*

16: em_est16:*

17: em_est17:*

18: em_est18:*

19: em_est19:*

20: gb_est1:*

21: gb_est2:*

22: gb_est3:*

23: gb_est4:*

24: gb_est5:*

25: gb_est6:*

26: gb_est7:*

27: gb_est8:*

28: gb_est9:*

29: gb_est10:*

30: gb_est11:*

31: gb_est12:*

32: gb_est13:*

33: gb_est14:*

34: gb_est15:*

35: gb_est16:*

36: gb_est17:*

37: gb_est18:*

38: gb_est19:*

39: gb_est20:*

40: gb_est21:*

41: gb_est22:*

42: gb_est23:*

43: gb_est24:*

44: gb_est25:*

45: gb_est26:*

46: gb_est27:*

47: gb_est28:*

48: gb_est29:*

49: gb_est30:*

50: gb_est31:*

51: gb_est32:*

52: em_est20:*

53: em_est21:*

54: em_est22:*

55: em_est23:*

56: em_est24:*

57: em_est25:*

58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	100.0	252	32	AA331546	AA331546 EST35390
2	17.8	77.4	478	43	AI102374	AI102374 EST211663
3	17.8	77.4	536	45	AA874982	AA874982 UT-R-E0-C
4	17.8	77.4	358	45	AI009082	AI009082 EST203533
5	17.8	77.4	497	48	AI547507	AI547507 UT-R-C3-S
6	17.8	77.4	436	48	AI576472	AI576472 UT-R-G0-U
7	17.2	74.8	376	30	AA196181	AA196181 zP96a06.s
8	17.2	74.8	316	31	AA325642	AA325642 EST28660
9	16.8	73.0	416	28	AA105882	AA105882 m184h07.r
10	16.8	73.0	465	29	AA164056	AA164056 m67506.r
11	16.8	73.0	376	30	AA272905	AA272905 v839d01.r
12	16.8	73.0	398	30	AA273650	AA273650 v832c09.r
13	16.8	73.0	491	31	AA289192	AA289192 v833c01.r
14	16.8	73.0	447	41	AU017898	AU017898 AU017898
15	16.8	73.0	91	44	AI133597	AI133597 EST051 MO
16	16.8	73.0	447	44	AI332915	AI332915 m67506.y
17	16.8	73.0	412	44	AI332858	AI332858 m67506.x
18	16.8	73.0	287	46	AI448486	AI448486 mp63f07.x
19	16.8	73.0	439	47	AI462305	AI462305 ub62b06.x
20	16.8	73.0	216	49	AV007110	AV007110 AV007110
21	16.8	73.0	403	49	AV008847	AV008847 AV008847
22	16.6	72.2	539	21	T83189	T83189 v441902.r1
23	16.6	72.2	355	22	R34570	R34570 yh85812.s1
24	16.4	71.3	482	34	AA530221	AA530221 v138a08.r
25	16.4	71.3	265	36	AA638506	AA638506 v054b01.r
26	16.4	71.3	474	42	AU023168	AU023168 AU023168
27	16.4	71.3	633	48	AI601278	AI601278 fC09a05.x
28	16.4	71.3	191	49	AV006692	AV006692 AV006692
29	16.4	71.3	180	50	AV030936	AV030936 AV030936
30	16.2	70.4	340	25	N43682	N43682 SWICAI1163S
31	16.2	70.4	651	29	AA130056	AA130056 z62e11.r
32	16.2	70.4	473	39	AA880834	AA880834 vx46d07.r
33	16.2	70.4	555	44	AI294344	AI294344 LP07690.5
34	16.2	70.4	480	49	AI638891	AI638891 AEMTAR20
35	16.2	70.4	662	51	AI721941	AI721941 fC25d12.x
36	16	69.6	326	46	AI425258	AI425258 vv74d05.y
37	16	69.6	412	46	AI451931	AI451931 mp74f04.x
38	16	69.6	468	47	AI530531	AI530531 mp74f04.y
39	15.8	68.7	185	20	Z41551	Z41551 HSCZXF122.N
40	15.8	68.7	339	24	H85019	H85019 y888e04.s1
41	15.8	68.7	549	28	AA121732	AA121732 z880n05.s
42	15.8	68.7	508	29	AA179551	AA179551 zp49b12.s
43	15.8	68.7	302	33	AA409963	AA409963 EST01904
44	15.8	68.7	435	40	AA938671	AA938671 OC08h07.s
45	15.8	68.7	341	41	AI028020	AI028020 cy73b03.x

ALIGNMENTS

RESULT 1
AA331546/c 252 bp mRNA
LOCUS AA331546
DEFINITION EST35390 Embryo, 8 week I Homo sapiens cDNA 5' end similar to
phosphatidylinositol 3-kinase, alpha catalytic subunit, mRNA
ACCESSION AA331546
NTD 91983787

```

VERSION AA311546.1 GI:1983787
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 252)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fullmer,R.A.,
Bull,C.J., Lee,N.H., Kirness,E.F., Weinstock,K.G., Gockayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man'wal,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glocke,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Metrick,M.,
Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Uteback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A.,
He,M.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Hasetline,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)
JOURNAL 96026280
MEDLINE
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404766.
Other ESTs: THC168479
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavage@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
4seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
source
1..252
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):133163"
/db_xref="taxon:9606"
/clone_1lb="Embryo, 8 week 1"
/dev_stage="Embryo, 8 wks"
/note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 78 a 46 c 53 g 75 t
ORIGIN
Query Match 100.0%; Score 23; DB 32; Length 252;
Best Local Similarity 100.0%; Pred. No. 0.083;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AATTCACACTGGCATGCCGAT 23
Db 119 AATTCACACTGGCATGCCGAT 97
RESULT 2
AII02374 478 bp mRNA EST 31-JAN-1999
LOCUS ESI21163 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
DEFINITION BBRC71.3' end, mRNA sequence.
ACCESSION AII02374
NID 93707163
VERSION AII02374.1 GI:3707163
KEYWORDS EST.
SOURCE Rattus sp.

```

```

ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 478)
AUTHORS Lee,N.H., Glocke,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
Unpublished (1998)
JOURNAL On Aug 21, 1998 this sequence version replaced.
COMMENT
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
Location/Qualifiers
source
1..478
/organism="Rattus sp."
/db_xref="ATCC (Inhost):2029673"
/db_xref="taxon:10118"
/clone_1lb="Normalized rat brain, Bento Soares"
/note="Organ: brain; Vector: pRT3pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 154 a 96 c 99 g 129 t
ORIGIN
Query Match 77.4%; Score 17.8; DB 43; Length 478;
Best Local Similarity 90.5%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 3 TTCACACTGGCATGCCGAT 23
Db 105 TTCACACTGGCATGCCGAT 125
RESULT 3
AA874982 536 bp mRNA EST 05-FEB-1999
LOCUS UT-R-EO-clf-c-06-0-UT.s1 UT-R-EO Rattus norvegicus cDNA clone
DEFINITION UT-R-EO-clf-c-06-0-UT 3' similar to g111669534[dbj]D67015D67015
Mouse mRNA for scg, complete cds, mRNA sequence.
ACCESSION AA874982
NID 94230862
VERSION AA874982.1 GI:4230862
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 536)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 97044477
MEDLINE
COMMENT On Mar 20, 1998 this sequence version replaced gi:2979930.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult 8-Day-Embryo library. cDNA library preparation: M. Fatima

```

Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics The following repetitive elements were found in this cDNA sequence: 1-35, >AL-rich#Low-complexity

Seq primer: M13 Forward.
Location/Qualifiers

1. 536
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/map="x"
/clone="UI-R-E0-cf-c-06-0-UI"
/clone_1lb="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: NotI; Site_2: EcoRI. This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."
BASE COUNT 169 a 105 c 108 g 154 t
ORIGIN

Query Match 77.4%; Score 17.8; DB 45; Length 536;
Best Local Similarity 90.5%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 TTCACACACTGGCATGCCGAT 23
||||| ||||||| ||
DB 122 TTCACACACTGGCATGCCAAT 142

RESULT 4
LOCUS A1009082 358 bp mRNA EST 08-JAN-1999
DEFINITION EST203533 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
REMARKS 3' end, mRNA sequence.
ACCESSION A1009082
NID 94132989
VERSION A1009082.1 GI:4132989
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 358)
AUTHORS Lee,N.H., Glodex,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat Gene Index
JOURNAL Unpublished (1998)
COMMENT On Jun 15, 1998 this sequence version replaced gi.3222914.

Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1. 358
/organism="Rattus sp."
/db_xref="taxon:10116"
/map="x" p12-q11
/clone="REMB132"
/clone_1lb="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pRT3D-Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT 97 a 77 c 69 g 115 t
ORIGIN

Query Match 77.4%; Score 17.8; DB 45; Length 358;
Best Local Similarity 90.5%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 TTCACACACTGGCATGCCGAT 23
||||| ||||||| ||
DB 257 TTCACACACTGGCATGCCAAT 237

RESULT 5
LOCUS A1547507 497 bp mRNA EST 22-MAR-1999
DEFINITION UI-R-C3-s1-g-08-0-UI-s1 UI-R-C3 Rattus norvegicus cDNA clone
UI-R-C3-s1-g-08-0-UI 3', mRNA sequence.
ACCESSION A1547507
NID 94464995
VERSION A1547507.1 GI:4464995
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
AUTHORS 1 (bases 1 to 497)
TITLE Ronaldo,M.F., Lennon,G. and Soares,M.B.
JOURNAL Normalization and subtraction: two approaches to facilitate gene
MEDLINE discovery
Genome Res. 6 (9), 791-806 (1996)
COMMENT 97044477
On Jun 5, 1998 this sequence version replaced gi.3187163.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msquares@iue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward.

FEATURES
source
Location/Qualifiers

1. 497
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/map="9 q22-q33; 898A03; 1; 1p32.1-1p33; 19"
/clone="UI-R-C3-s1-g-08-0-UI"
/clone_1lb="UI-R-C3"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI. The UI-R-C3 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, and 8, 12 and 18-day embryos, after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: UI-R-C3, UI-R-C2p, UI-R-C1, UI-R-C0, UI-R-A1, UI-R-E1. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C3) was constructed as follows: PCR amplified cDNA inserts from UI-R-C2p clones from which 3' ESTs had

been derived was used as a driver in a hybridization with the UI-R-C2p library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C3 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6:791-806, 1996).

Research 6:791-806, 1996)

BASE COUNT 155 a 100 c 99 g 143 t

Query Match 77.4%; Score 17.8; DB 48; Length 497;
Best Local Similarity 90.5%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 TTCACACTGGCAGCCGAT 23
||||| ||||||| ||
Db 119 TTCACACTGGCAGCCGAT 139

RESULT 6
A1576472 436 bp mRNA EST 05-APR-1999
LOCUS UI-R-GO-us-e-01-0-UI.s1 UI-R-GO Rattus norvegicus cDNA clone
DEFINITION UI-R-GO-us-e-01-0-UI 3', mRNA sequence.
ACCESSION A1576472
NID 94560848
VERSION A1576472.1 GI:4560848
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 436)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
On Jun 5, 1998 this sequence version replaced gi:3188268.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.wieg.uiowa.edu

The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonaldi poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the normalized ganglia library cDNA library preparation. M.B. Soares
Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 forward.
Location/Qualifiers

FEATURES

source
1..436
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-GO-us-e-01-0-UI"
/clone_1ib="UI-R-GO"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI. The UI-R-GO library is a normalized library constructed from a mixture of rat tissues (nodose ganglia, dorsal root ganglia, and trigeminal ganglia). The tag is a string of

6 nucleotides present between the Not I site and the oligo-dt track. The library was constructed as described by Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996."

BASE COUNT 140 a 79 c 93 g 124 t

OY 3 TTCACACTGGCAGCCGAT 23
||||| ||||||| ||
Db 119 TTCACACTGGCAGCCGAT 139

RESULT 7
AA196181/c 376 bp mRNA EST 10-MAR-1998
LOCUS zp96a06.s1 Stratiagene muscle 937209 Homo sapiens cDNA clone
DEFINITION IMAGE:628018 3', mRNA sequence.
ACCESSION AA196181
NID 91791763
VERSION AA196181.1 GI:1791763
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 376)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kudaba, T., Lacy, M., Le, N., Lennon, G., Matra, M.,
Martin, J., Moore, B., Scheinberg, R., Steptoe, M., Tan, F.,
Thaising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1397572.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 1118 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 213.
Location/Qualifiers

FEATURES

source
1..376
/organism="Homo sapiens"
/db_xref="GDB:5048252"
/db_xref="taxon:9606"
/map="17q21"
/clone="IMAGE:628018"
/clone_1ib="Stratiagene muscle 937209"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="SOLR (Kanamycin resistant)"
/note="Organ: skeletal muscle; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
Primer: Oligo dt. Skeletal muscle from patient with
malignant hyperthermia. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG
3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'."

BASE COUNT 97 a 78 c 77 g 124 t

Query Match 74.8%; Score 17.2; DB 30; Length 376;
Best Local Similarity 86.4%; Pred. No. 66;


```

ACCESSION      IMAGE:635002.5', mRNA sequence.
NID            AA164056
VERSION        91740009
KEYWORDS       AA164056.1 GI:1740009
SOURCE         EST.
ORGANISM       house mouse.
AUTHORS        Mus musculus.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 465)
AUTHORS        Marra,M., Hillier,L., Allen,M., Bowles,M., Dierlich,N., Dubuque,T.,
               Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
               Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
               Thaising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
               Waterston,R.
TITLE          The Mashu-HHMI Mouse EST Project
JOURNAL        Unpublished (1996)
COMMENT        On Jan 25, 1995 this sequence version replaced g1:637825.

FEATURES             location/qualifiers
     source           1..465
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /map="22"
                     /clone="IMAGE:635002"
                     /clone_1lb="Soares mouse lymph node NBMUN"
                     /sex="male"
                     /tissue_type="lymph node"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: lymph node; Vector: pTZ19D-Pac (Pharmacia)
                     with a modified polylinker; Site.1: Not I; Site.2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5' TGTTACCATGTGAAGTGCGGAGCCGCCGCATCTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pTZ19 vector. RNA
                     provided by Dr. Bertrand Jordan, library constructed and
                     normalized by Bento Soares and M.Fatima Ronaldo."

BASE COUNT      *    140 a          94 c          82 g          149 t
ORIGIN
Query Match      73.0%; Score 16.8; DB 29; Length 465;
Best Local Similarity 90.0%; Pred. No.1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY              3 TTCACACTGGCATGCCGA 22
                ||||| ||||||||| |
Db              329 TTCACAGACTGGCATGCCAA 310

RESULT 11
LOCUS          AA272905          376 bp      mRNA      EST      26-MAR-1997
DEFINITION     va39d01.r1 Soares mouse JUNE12.5 Mus musculus cDNA clone
ACCESSION      AA272905
VERSION        91907497
GI             91907497

```

```

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
Oy
Db
RESULT 12
LOCUS
DEFINITION
ACCESSION
MID
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
house mouse.
mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 376)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kuchta,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMNI Mouse EST Project
On Sep 12, 1996 this sequence version replaced gi:1394067.
Contact: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL: contact the
IMAGS Consortium (info@image.lnl.gov) for further information.
MGI:450777
Seq primer: -28mj rev2 ET from Amersham
High quality sequence stop: 361.
Location/Qualifiers
1. 376
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:733729"
/clone_1ib="Scopes mouse 3NNE12 5"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
/note="Organ: Whole fetus; Vector: pT773D-Pec (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTTCCACATCTGGAAGTGGAGCGCGCCGCTATTTTCTTTTCTTTT
3'], on total mouse RNA [provided by Minoru Ko, Wayne
State Univ.]; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
128 a 77 c 78 g 93 t
2 ATTGCACACTGCATGCCG 21
|||||
81 ATTGCACACTGCATGCTG 100
Query Match 73.0%; Score 16.8; DB 30; Length 376;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0
Oy 2 ATTGCACACTGCATGCCG 21
|||||
Db 81 ATTGCACACTGCATGCTG 100
RESULT 12
LOCUS AA273650 398 bp mRNA EST 28-MAR-1997
DEFINITION V092C09.r1 Soares mouse 3NBMS Mus musculus cDNA clone IMAGE:764464
ACCESSION AA273650
MID 91912717
VERSION AA273650.1 GI:1912717
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```


FEATURES
source
1..638
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/cell_type="oocyte"
107..583
/gene="ubc9"
107..583
/gene="ubc9"
/note="ubc9"
/codon_start=1
/product="E2 ubiquitin conjugating enzyme"
/protein_id="AA57736.1"
/db_xref="GI:2078331"
/translation="MSGIALSLAQRKARNDHPGFVAVPKNDGIMNIMWCA
IPKKGTPWGGGLFKLMLEFDYPPSPKCEPPLFHNVPSTGVCISLEEDK
MRPATIKQLLQIOLNEPNIODPAQAEAYITICQNRVEYKRVRAQAKKAPAS"

BASE COUNT 214 a 131 c 147 g 146 t

ORIGIN

Query Match 100.0%; Score 35; DB 4; Length 638;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTGTGTGTGTGT 35
|||||

DB 638 GACTCGAGTCGACATCGATTTTGTGTGTGT 604
|||||

RESULT 2
A24382
LOCUS A24382 35 bp DNA PAT 21-MAR-1995
DEFINITION d517 adaptor primer.
ACCESSION A24382
NID 9904477
VERSION A24382.1 GI:904477
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 35)
AUTHORS Sarmientos, P.; De Taxis du Poet, P.; Nittl, G. and Scacheri, E.
TITLE Anti-thrombin polypeptides
JOURNAL Patent: EP 0501821 A 22 02-SEP-1992;
FARMITALIA CARLO ERBA S.R.L.
FEATURES
source
1..35
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 5 a 5 c 5 g 20 t

ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTGTGTGTGT 35
|||||

DB 1 GACTCGAGTCGACATCGATTTTGTGTGTGT 35
|||||

RESULT 3
A27645
LOCUS A27645 35 bp DNA PAT 04-JUN-1995
DEFINITION Hybrid dT17-adaptor primer.
ACCESSION A27645
NID 91248481
VERSION A27645.1 GI:1248481
KEYWORDS
SOURCE synthetic construct.

ORGANISM synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 35)
AUTHORS
TITLE POLYPHENOL OXIDASE GENES
JOURNAL Patent: WO 9302195-A 1 04-FEB-1993;
FEATURES
source
1..35
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 5 a 5 c 5 g 20 t

ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTGTGTGTGT 35
|||||

DB 1 GACTCGAGTCGACATCGATTTTGTGTGTGT 35
|||||

RESULT 4
A37244
LOCUS A37244 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 13 from Patent WO9403609.
ACCESSION A37244
NID 92294355
VERSION A37244.1 GI:2294355
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Parker, P.J.; Goode, N.T.; Nurse, P.M. and Waterfield, M.D.
TITLE EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 13 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
source
1..35
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 5 a 5 c 5 g 20 t

ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTGTGTGTGT 35
|||||

DB 1 GACTCGAGTCGACATCGATTTTGTGTGTGT 35
|||||

RESULT 5
A40373
LOCUS A40373 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 29 from Patent WO9425489.
ACCESSION A40373
NID 92296422
VERSION A40373.1 GI:2296422
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Mohapatra, S.S. and Sehon, A.H.
TITLE DOWNREGULATION OF ALLERGEN-SPECIFIC IMMUNE RESPONSE
JOURNAL Patent: WO 9425489-A 29 10-NOV-1994;

UNIV MANITOBA (CA)
COMMENT Other publication AU 6674094 941121.
FEATURES
source
1. .35
/db_xref="taxon:32644"
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 35; DB 5; Length 35;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
Db 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 6
LOCUS A42335 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 7 from Patent WO9502057.
ACCESSION A42335
NID 92297812
VERSION A42335.1 GI:2297812
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Gusterson,B.A., Crompton,M.R., Mitchell,P.J., Barker,K.T.,
TITLE Kamelati,T., Page,M.J. and Spence,P.
JOURNAL PROTEIN TYROSINE KINASE AND LIGANDS THEREOF
COMMENT Patent: WO 9502057-A 7 19-JAN-1995;
FEATURES Other publication AU 7080994 950206.
source
1. .35
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 35; DB 5; Length 35;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
Db 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 7
LOCUS A42384 35 bp DNA PAT 06-MAR-1997
DEFINITION Sequence 7 from Patent WO9502187.
ACCESSION A42384
NID 92297858
VERSION A42384.1 GI:2297858
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Gusterson,B.A., Crompton,M.R., Mitchell,P.J., Barker,K.T.,
TITLE Kamelati,T., Page,M.J. and Spence,P.
JOURNAL CELL GROWTH FACTOR RECEPTORS
COMMENT Patent: WO 9502187-A 7 19-JAN-1995;
FEATURES Other publication AU 7081094 950206.
source
1. .40
Location/Qualifiers

source
1. .35
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 35; DB 5; Length 35;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
Db 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 8
LOCUS A46467 35 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 5 from Patent WO9526402.
ACCESSION A46467
NID 92300644
VERSION A46467.1 GI:2300644
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Knox,D.P., Smith,S.K., Smith,W.D., Redmond,D. and Murray,J.
TITLE VACCINES AGAINST HELMINTHIC PARASITES
JOURNAL Patent: WO 9526402-A 5 05-OCT-1995;
COMMENT MALINCKRODT VETERINARY INC (US)
Other publication ZA 9502370 951215
Other publication CA 2182178 951005
Other publication AU 1956495 951017.
FEATURES Location/Qualifiers
source
1. .35
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 35; DB 5; Length 35;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
Db 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 9
LOCUS A59198 40 bp DNA PAT 06-MAR-1998
DEFINITION Sequence 7 from Patent WO9704108.
ACCESSION A59198
NID 93714579
VERSION A59198.1 GI:3714579
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 40)
AUTHORS Schuster,E., Spiessler,B., Titzel,K., Gottschalk,M., Khan,N.Q.,
TITLE Wolf,S. and Plainer,H.
JOURNAL LECITINE AMINOPEPTIDASES PRODUCED RECOMBINANTLY FROM ASPERGILLUS
COMMENT Patent: WO 9704108-A 7 06-FEB-1997;
FEATURES ROEHM GMBH (DE)
Other publication DE 19526485 970123.
source
1. .40
Location/Qualifiers

LOCUS 128284 35 bp DNA PAT 30-OCT-1996

DEFINITION Sequence 13 from patent US 5569830.

ACCESSION 128284

NID 91819060

VERSION 128284.1 GI:1819060

KEYWORDS

SOURCE Unknown.

ORGANISM

REFERENCE Unclassified.

1 (bases 1 to 35)

AUTHORS Bennett, A., Labavitch, J. M., Powell, A. and Stoltz, H.

TITLE Plant inhibitors of fungal polygalacturonases and their use to

control fungal disease

Patent: US 5569830-A 13 29-OCT-1996;

Location/Qualifiers

FEATURES

1. .35

source /organism="unknown"

BASE COUNT 5 a 5 c 5 g 20 t

ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.095;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

Db 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

RESULT 15

134242

LOCUS 134242 35 bp DNA PAT 30-JAN-1997

DEFINITION Sequence 19 from patent US 5597569.

ACCESSION 134242

NID 91825033

VERSION 134242.1 GI:1825033

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE Unclassified.

1 (bases 1 to 35)

AUTHORS Siegfall, C. B., Gawlak, S. L. and Marguardt, H.

TITLE Bryodin 2 a ribosome-inactivating protein isolated from the plant

Bryonia dioica

Patent: US 5597569-A 19 28-JAN-1997;

Location/Qualifiers

FEATURES

1. .35

source /organism="unknown"

BASE COUNT 5 a 5 c 5 g 20 t

ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.095;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

Db 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

134242

LOCUS 134242 35 bp DNA PAT 30-JAN-1997

DEFINITION Sequence 19 from patent US 5597569.

ACCESSION 134242

NID 91825033

VERSION 134242.1 GI:1825033

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE Unclassified.

1 (bases 1 to 35)

AUTHORS Siegfall, C. B., Gawlak, S. L. and Marguardt, H.

TITLE Bryodin 2 a ribosome-inactivating protein isolated from the plant

Bryonia dioica

Patent: US 5597569-A 19 28-JAN-1997;

Location/Qualifiers

FEATURES

1. .35

source /organism="unknown"

Search completed: September 13, 1999, 15:55:19

Job time: 4547 sec

THIS PAGE BLANK (USPTO)

KM polymerase chain reaction; amplify: Haemonchus contortus; ss.
 OS Synthetic.
 PN WO9323542-A.
 PD 25-NOV-1993.
 PE 07-MAY-1993; G00943.
 PR 08-MAY-1992; GB-009993.
 (AGRI-) AGRIC 6 FOOD RES COUNCIL.
 PI Graham M, Knox DP, Munn EA, Newton SE, Oliver JJ;
 PI Smith TS;
 DR WPI; 93-386574/48.
 PT New DNA encoding amino-peptidase from Haemonchus contortus and
 PT derived peptide(s) - useful in protective vaccines, for new
 PT vectors, transformed cells and oligosaccharide, for incorporation
 PT in virus or microbe
 PI Example; Page 32; 137pp; English.
 CC The sequences given in 052503-11 are primers which were used to
 CC amplify cDNA clones of the helminth aminopeptidase genes H11-1, -2
 CC and -3. The amplified sequences encode H110D which is a protein
 CC doublet which shows homology to a family of integral membrane amino-
 CC peptidases. The differences between the amplified clones can be
 CC attributed to different mRNAs of the multigene family, and also to
 CC different variants of the H110D-encoding sequence being present at
 CC different stages of the life cycle, or in strains differing in
 CC geographical origin. Antigenic fragments of the aminopeptidases
 CC encoded by the H11 genes may be used in vaccines to stimulate immune
 CC response against helminth parasites in humans or other animals. These
 CC DNA sequences may be incorporated into a virus or microbe and used in
 CC a similar manner.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;
 Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 GACTCGAGTCGACATCGATTTT 35
 1 GACTCGAGTCGACATCGATTTT 35
 RESULT 3
 ID 059015 standard; DNA: 35 BP.
 AC 059015:
 DT 12-APR-1994 (first entry)
 DE RACE PCR adaptor-dt.
 KW phosphonositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphonositides;
 KW ss; amplification; p110.
 OS Synthetic.
 PN WO9321328-A.
 PD 28-OCT-1993.
 PE 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDM-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parnerp J, Volinia S, Waterfield MD;
 DR WPI; 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphonositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1; Page 39; 146pp; English.
 CC An SGBAF-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with pSV3neo. Total RNA was
 CC isolated from the SGBAF-1 and a cDNA library constructed. RACE PCR
 CC was performed by synthesizing first strand cDNA from random hexamers
 CC on SGBAF-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-
 CC dt as primers. Products were fractionated using an agarose gel. The
 CC DNA was isolated from the gel and subjected to PCR using oligo 2280 and
 CC adaptor as primers. A product of 350 bp was further sequenced.
 CC See also 051155-6, 059012-23 and 057522-3.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 GACTCGAGTCGACATCGATTTT 35
 1 GACTCGAGTCGACATCGATTTT 35
 RESULT 4
 ID 057021 standard; DNA: 35 BP.
 AC 057021:
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase adaptor dt primer.
 KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schistosomacromyces pombe; nmt promoter; thiamine;
 KW ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Synthetic.
 PN WO9403609-A.
 PD 17-FEB-1994.
 PE 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-016554.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PUJ, Waterfield MD;
 DR WPI; 94-065697/08.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Example 1; Page 15; 71pp; English.
 CC The sequences given in 057020-23 are primers which were used in the
 CC isolation of the phosphatidylinositol (Ptdins) 3-kinase cDNA. The
 CC amplified sequence was placed under the regulatory control of the nmt
 CC promoter. These sequences were transformed into Schistosomacromyces
 CC pombe cells in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the Ptdins
 CC plasmids grow as the parental strain. In the absence of thiamine the
 CC nmt promoter functions and the Ptdins is induced. Ptdins activity is
 CC substantially increased under these conditions. Cells containing
 CC constructs such as this, are useful in assays for detecting compounds
 CC involved in cell growth regulation. It is also used as the basis for
 CC detecting compounds for treating cancers and the formation of blood
 CC vessel plaques.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;
 Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 GACTCGAGTCGACATCGATTTT 35
 1 GACTCGAGTCGACATCGATTTT 35
 RESULT 5
 ID 071021 standard; DNA: 35 BP.
 AC 071021:
 DT 26-SEP-1994 (first entry)
 DE Primer for amplifying peach polygalacturonase coding sequence.
 KW Ripening; melting; pectins; breeding; Prunus; Rosaceae;
 KW endopolygalacturonase; EndoPG; polygalacturonase; PG; ss.
 OS Synthetic.
 PN WO9405795-A.
 PD 17-MAR-1994.
 PE 27-AUG-1993; AU0442.
 PR 27-AUG-1992; AU-004370.
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PA (PACT-) PACIFIC SEEDS PVT LTD.
 PI Brady CJ, Lee E, Lester DR, Orr G, Speirs J;

DR WPI: 94-101199/12.
 PT DNA encoding peach polygalacturonase (Pg) and production of
 PT recombinant peach Pg in rosaceae plants for production of
 PT plants with 'melting' or 'non-melting' phenotype
 PS Disclosure; Page 15; 51pp; English.
 CC During early ripening of peaches, tissue firmness decreases
 CC slowly and progressively. Towards the end of ripening, loss of
 CC tissue firmness is rapid. This second stage of softening is called
 CC the "melting" stage. Fruit of peach varieties used for canning do
 CC not have a "melting" phase of softening. Ripe fruit remain
 CC relatively firm and maintain their shape throughout processing.
 CC Fruit of "melting" varieties show an increase in activity of
 CC endopolygalacturonase (PnPG) during ripening. Polygalacturonase
 CC (Pg) is believed to contribute to fruit softening through its action
 CC on intercellular and cell wall pectins. This primer was used to
 CC synthesize a first strand cDNA molecule for further amplification
 CC by PCR. See 056765-67.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35
 DB 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 6
 063873
 ID 063873 standard; cDNA; 35 BP.
 AC 063873:
 DE T cell protein CD4 adaptor-dt17 primer.
 KW Cat; feline immunodeficiency virus; CD4 positive T lymphocyte;
 KW cellular receptor; T cell protein; polymerase chain reaction; PCR; ss.
 OS Synthetic.
 PN FR2696188-A.
 PD 01-APR-1994.
 PR 25-SEP-1992; 011496.
 PR 25-SEP-1992; FR-011496.
 PA (UWPA-) UNIV CURIE PARIS VI P 6 M.
 PI De Parseval A, Klatzmann D, Salmon P, Sonigo P;
 DR WPI: 94-128285/16.
 PL Feline CD4 protein and corresp. DNA - for study of feline
 PT immunodeficiency virus and for therapy
 PS Disclosure; Page 5; 25pp; French.
 CC The primers 063871 and 063872 were used to obtain cDNA coding for
 CC feline CD4 protein by polymerase chain reaction. For rapid
 CC amplification of the ends of the cDNA, a hybrid adaptor-oligo(dt)
 CC primer was also used. Cat lymph node cDNA was amplified and the
 CC full-length CD4 coding sequence 063870 was obtained from several
 CC partial clones.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35
 DB 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 7
 065594
 ID 065594 standard; DNA; 35 BP.
 AC 065594:
 DE HER4 primer XSC17.
 KW Human epidermal growth factor receptor; HER; EGF; tyrosine kinase;

KW cancer; neuronal tissue; muscle tissue; neoplasm;
 KW carcinoma; primer; probe; PCR; ss.
 OS Synthetic.
 PN EP-599274-A.
 PD 01-JUN-1994.
 PF 23-NOV-1993; 118837.
 PR 24-NOV-1997; US-981165.
 PA (BRIN) BRISTOL-MYERS SQUIBB CO.
 PI Culoscou J, Plowman GD, Shoyab M;
 DR WPI: 94-169599/21.
 PT New recombinant nucleic acid expressing HER4 - a new receptor
 PT tyrosine kinase expressed in some cancer cells, and related
 PT vectors, antibodies, ligands etc, for diagnosis and treatment of
 PT cancers.
 PS Example; Page 16; 104pp; English.
 CC The primers given in 065594-6 were used in the semi-quantitative
 CC PCR detection of HER4.
 CC HER4 is the fourth member of the EGFR-family of tyrosine kinases and
 CC is expressed in some human cancers and in some tissues of neuronal
 CC or muscle origin. HER4 polynucleotides, opt. labelled, are useful
 CC in assays (e.g. of HER4 mRNA to detect certain neoplasms, esp.
 CC breast carcinoma) and as primers in PCR or as probes.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35
 DB 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 8
 078807
 ID 078807 standard; cDNA; 35 BP.
 AC 078807:
 DE 12-JUN-1995 (first entry)
 DT TCR alpha-chain dt17 adaptor.
 KW Allergen; immunotherapy; T-cell receptor; TCR; CD8;
 KW complementarity determining region-3; immunosuppressive;
 KW antibody engineering; suppressor T-lymphocyte; Ts; PCR;
 KW polymerase chain reaction; amplification; adaptor; ss.
 OS Synthetic.
 PN WO9425489-A.
 PD 10-NOV-1994.
 PF 26-APR-1994; CA0228.
 PR 26-APR-1995; GB-008381.
 PA (UYMA-) UNIV MANITOBA.
 PI Mohapatra SS, Sehon AH;
 DR WPI: 94-358193/44.
 PT Synthetic peptides for effecting immuno-therapy - have an amino
 PT acid sequence comprising a portion of the CDR3 region of a T-cell
 PT receptor
 PS Disclosure; Page 43; 78pp; English.
 CC Antibody production against allergenic antigen (Ag) is specifically
 CC suppressed by treatment with the junctional segment of the alpha
 CC and/or beta chain of the Ag receptor of suppressor T (Ts) cells
 CC (TCR) induced by tolerogenic Ag-Peg conjugates. cDNAs encoding
 CC the alpha and beta chains of TCRs of cloned Ts cells specific for
 CC ovalbumin and human monoclonal (myeloma) IgG were produced by PCR
 CC using the primers given in 078798-805 and the adaptors given in
 CC 078806-07.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35
 DB 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 9

ID 084786 standard; DNA; 35 BP.

AC 084786

DT 17-AUG-1995 (first entry)

DE PTK2 reverse-transcription primer.

KW protein-tyrosine-kinase; PTK; discoidin domain receptor; cancer;

KM breast tumor; mamma carcinoma; diagnosis; prognosis; therapy; ss.

OS Synthetic.

PN WO9502187-A.

PD 19-JAN-1995.

PF 08-JUL-1994; G01480.

PR 09-JUL-1993; GB-014271.

PA (CANC-) CANCER RES INST.

PI (WELL) WELLCOME FOUND LTD.

PI Barker KT, Crompton MR, Gusterson BA, Martindale JE;

PI Mitchell PJ, Spence P;

DR WPI: 95-066991/09.

PT Method for screening substances, using protein tyrosine kinase -

PT for potential utility as therapeutic agents for cancer

PS Disclosure; Page 34; 51pp; English.

CC CDNA derived from tumor metastatic tissue was amplified using

CC primers (given in 084783-84) based on sequences (R71101, R71103)

CC associated with protein-tyrosine-kinases (PTK). Novel PTK2 was

CC identified in an isolated subclone. The 3' sequence of PTK2 was

CC obtained by reverse transcription (using the primer of 084786) and

CC PCR amplification (primers 084787-88) of RNA of human breast

CC carcinoma cell line MDA MG 468. The partial DNA sequence of PTK2

CC is given in 084782.

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 10

ID 090167 standard; DNA; 35 BP.

AC 090167;

DT 01-NOV-1995 (first entry)

DE Pollen allergen gene primer 8.

KW Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;

KM sensitizier; Cryptomeria japonica; polymerase chain reaction;

KW PCR; primer; ss.

OS Synthetic.

PN EP-655500-A.

PD 31-MAY-1995.

PF 03-NOV-1994; 308117.

PR 05-NOV-1993; JP-289151.

PR 20-DEC-1993; JP-344596.

PR 27-DEC-1993; JP-346814.

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PI Kurimoto M, Namba M, Torisoe K;

PI WPI: 95-195588/26.

DR New Japanese cedar pollen allergen polypeptide - and DNA coding for it,

PT useful for treatment and diagnosis of cedar pollen allergy

PS Disclosure; Page 11; 41pp; English.

CC The primers given in 090166-69, based on a portion of cDNA clone

CC SC09 or on previously isolated sequences, were used for the PCR

CC amplification of cDNA derived from Japanese cedar pollen, to obtain

CC clone SC50 bearing a partial sequence (nt 1-198 of the sequence

CC given in 090156) of the pollen allergen gene.

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 11

ID 089878 standard; DNA; 35 BP.

AC 089878;

DT 31-JAN-1997 (Revised)

DT 25-OCT-1995 (first entry)

DE PCR primer for cloning bryodin-2.

KW Ribosome inactivating protein; cell death; proliferation; tumour;

KM cancer; virus; malaria; trypanosomiasis; inflammation; autoimmune;

KM disease; purging; bone marrow; ss.

OS Synthetic.

PN WO9511977-A.

PD 04-MAY-1995.

PF 25-OCT-1994; U12382.

PR 25-OCT-1993; US-141891.

PR 20-OCT-1994; US-324301.

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Gawlak SL, Marguardt H, Siegal CB;

DR WPI: 95-178870/23.

PT New ribosome inactivating protein bryodin-2 and related

PT conjugates - for killing target cells, e.g. in treatment of

PT cancer, also related nucleic acid, vectors and transformed cells

PS Disclosure; Page 43; 81pp; English.

CC The sequence is that of a PCR primer used to isolate a ribosome

CC inactivating protein termed bryodin-2, from Bryonia dioica. Conjugates

CC of the protein with a ligand may be used to kill target cells (i.e. those

CC to which the ligand component binds specifically) or to inhibit

CC proliferation of tumour cells. Typical applications include the

CC treatment of cancer, viral infections, malaria, trypanosomiasis,

CC inflammatory or autoimmune diseases, including in vitro purging of e.g.

CC bone marrow.

CC See also 089874-86.

CC Updated 31-JAN-1997 due to incorrect cross reference numbers.

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 12

ID T05321 standard; DNA; 35 BP.

AC T05321;

DT 13-APR-1996 (first entry)

DE Oligonucleotide probe for human prostacyclin-synthase.

KW DNA probe; prostacyclin-synthase; PCR; polymerase chain reaction;

KM DNA primer; prostaglandin I2; circulatory disease; therapeutic;

KW diagnosis; gene therapy; ss.

OS Synthetic.

PN WO9530013-A1.

PD 09-NOV-1995.

PF 27-APR-1995; J00838.

PR 28-APR-1994; JP-114316.

PA (TANA/) TANABE T.

PI Tanabe T;

DR WPI: 95-393084/50.

PT Human prostacyclin synthase and DNA encoding it - useful in the

```

PT Investigation and treatment of diseases characterised by reduced
PT prostaglandin 12 production.
PS Disclosure; Page 34; 71pp; Japanese.
CC DNA primers (T05317-20; T05322; T05326-27) are used to screen human
CC genomic lung cell line W138 and human arterial endothelial cell
CC cDNA libraries for the isolation of a prostacyclin-synthase (PGIS)
CC coding sequence (see T05316). DNA probes (T05321 and T05323) are used
CC in the construction of plasmid pHPG1S1, encoding the complete PGIS
CC sequence. This plasmid was used to transfect human 293 cells for
CC PGIS peptide expression. DNA encoding human PGIS, vectors containing
CC 1 $\alpha$ , and PGIS itself, may be administered to patients to increase
CC prostaglandin 12 (PGI2) production to treat diseases characterized
CC by reduced PGI2 levels or by an imbalance between PGI2 and
CC thromboxane A2 levels, such as circulatory diseases (thrombosis,
CC angina pectoris, arteriosclerosis, myocardial infarction). The
CC cDNA and protein are also useful in disease diagnosis.
SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35
DB 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35

RESULT 13
OY 094244
ID 094244 standard; DNA: 35 BP.
AC 094244;
DT 14-MAY-1996 (first entry)
DE Cysteine proteinase active site antisense polyT primer.
KW Primer; amplify; polymerase chain reaction; PCR; Haemonchus contortus;
KW cysteine proteinase; DM.2; DM.3; DM.4; human; DM.4a; DM.5; antigen;
KW vaccine; helminth; parasite; DM.1; ruminant; gastro-intestinal tract; ss.
OS Synthetic.
PN W09526402-A1.
PD 05-OCT-1995.
PF 24-MAR-1995; G00065.
PR 25-MAR-1994; GB-005990.
PR 25-MAR-1994; GB-005925.
PA (MCM ) MALLINCKRODT VETERINARY INC.
PI PI Knox DP, Murray J, Redmond D, Smith SK, Smith WD;
DR WPI; 96-351332/45.
PT Protective helminth parasite antigen - used in vaccine directed
PT against parasitic nematodes of mammalian gastro-intestinal tract
PS e.g. Haemonchus contortus
PS Example 16; Fig. 15; 79pp; English.
CC The sequences given in 094240-45 are primers which are based on
CC previously published sequences derived from the canonical Haemonchus
CC contortus cysteine proteinase molecule. These primers were used in
CC the cloning of cDNA fragments from the cysteine proteinase gene,
CC such as DM.1, DM.2, DM.3, DM.4, DM.4a and DM.5 (see also 094246-51).
CC The amplified fragments may be expressed in a recombinant cell for
CC the production of antigens. These antigens may be used in the
CC preparation of a vaccine against helminth parasites in a human or
CC non-human animal.
SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35
DB 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35

RESULT 14
OY 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35
DB 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35

RESULT 15
OY 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35
DB 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35

Query Match 100.0%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35
DB 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35

RESULT 15
OY T10698/c
ID T10698 standard; cDNA to mRNA; 901 BP.
AC T10698;
DT 26-MAY-1996 (first entry)
DE Trypsin-like enzyme N-terminal DNA fragment from p19-33.
KW Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;
KW asthma; VIP; vasoactive intestinal peptide; Influenza virus;
KW protease; primer; PCR; amplification; ds.
OS Homo sapiens.
PN AU9527248-A.
PD 08-FEB-1996.
PF 31-JUL-1995; 027248.
PR 29-JUL-1994; JP-178607.
PA (TEIJ ) TEIJIN LTD.
PI Masuda K, Ogawa H, Suga T, Sugimoto Y, Takagi K;
PI Yamaoka K, Yamaoka S;
DR WPI; 96-117356/13.
PT Nucleic acid sequence encoding trypsin-like enzyme - which digests
PT fibrinogen, used as expectorant in treatment of respiratory
PT diseases, e.g. bronchial asthma
PS Example 9; Page 44-45; 65pp; English.
CC Example 9 describes the cloning of cDNA region encoding
CC trypsin-like enzyme isolated from cough phlegm.
CC Four oligonucleotides are used for PCR: TRX-1 (T10694), TRX-8
CC (T10695), TRX-10 (T10696) and TRX-11 (T10697).
CC TRX-1 corresponds to from A1 to A23 of the sequence given in
CC T10693, which is part of a cDNA encoding the trypsin-like
CC enzyme. TRX-8 corresponds to from G16 to T40 of T10693.

```


OM nucleic - nucleic search, using sw model

(without alignments)
40.740 Million cell updates/sec

Title: US-09-325-095-16

Sequence: 1 GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT 35

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database

	EST1*
1:	em_est1.*
2:	em_est2.*
3:	em_est3.*
4:	em_est4.*
5:	em_est5.*
6:	em_est6.*
7:	em_est7.*
8:	em_est8.*
9:	em_est9.*
10:	em_est10.*
11:	em_est11.*
12:	em_est12.*
13:	em_est13.*
14:	em_est14.*
15:	em_est15.*
16:	em_est16.*
17:	em_est17.*
18:	em_est18.*
19:	em_est19.*
20:	qb_est1.*
21:	qb_est2.*
22:	qb_est3.*
23:	qb_est4.*
24:	qb_est5.*
25:	qb_est6.*
26:	qb_est7.*
27:	qb_est8.*
28:	qb_est9.*
29:	qb_est10.*
30:	qb_est11.*
31:	qb_est12.*
32:	qb_est13.*
33:	qb_est14.*
34:	qb_est15.*
35:	qb_est16.*
36:	qb_est17.*
37:	qb_est18.*
38:	qb_est19.*
39:	qb_est20.*
40:	qb_est21.*
41:	qb_est22.*
42:	qb_est23.*
43:	qb_est24.*
44:	qb_est25.*
45:	qb_est26.*
46:	qb_est27.*
47:	qb_est28.*
48:	qb_est29.*
49:	qb_est30.*
50:	qb_est31.*
51:	qb_est32.*
52:	em_est40.*
53:	em_est41.*

```
54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	100.0	337	32	AA340759	EST46034
2	25.6	73.1	678	44	A1295887	A1295887 LP09649.5
3	25.4	72.6	380	44	A1312326	A1312326 ta07603.x
4	25.4	72.6	111	45	A1348883	A1348883 th05c11.x
5	25.4	72.6	909	48	A1568751	A1568751 th15609.x
6	25.4	72.6	466	51	A1717729	A1717729 tl-R-y0-a
7	24.8	70.9	305	23	L38011	L38011 BNAF0228E M
8	24.8	70.9	435	28	C18175	C18175 C18175 Humma
9	24.6	70.3	314	35	AA541358	AA541358 n177604.s
10	24.6	70.3	466	51	A1728205	A1728205 BNLGH15255
11	24.6	70.3	623	51	A1728285	A1728285 BNLGH15355
12	24.4	69.7	405	35	AA541329	AA541329 n176611.s
13	24.4	69.7	511	51	A1732323	A1732323 n135f10.x
14	24.4	69.7	517	51	A1734188	A1734188 n078612.y
15	24.2	69.1	570	29	AA145612	AA145612 ms07h10.t
16	24.2	69.1	514	38	AA151165	AA151165 ISMJ0847
17	24.2	69.1	656	44	A1322793	A1322793 ms07h10.y
18	24.2	69.1	468	47	A1517504	A1517504 GSK86493.5
19	24	68.6	168	43	A1223480	A1223480 qx27b01.x
20	24	68.6	149	43	A1223537	A1223537 qx28b03.x
21	24	68.6	146	43	A1224352	A1224352 qw99d09.x
22	24	68.6	64	43	A1224631	A1224631 qw97e03.x
23	24	68.6	92	44	A1246064	A1246064 qv52e10.x
24	24	68.6	143	44	A1250505	A1250505 qx33b05.x
25	24	68.6	137	44	A1250569	A1250569 qx31e07.x
26	24	68.6	267	44	A1251260	A1251260 qx36e05.x
27	24	68.6	215	44	A1251658	A1251658 qv32b04.x
28	24	68.6	90	44	A1252577	A1252577 qv25e08.x
29	24	68.6	239	44	A1252739	A1252739 qv26b09.x
30	24	68.6	360	44	A1254267	A1254267 qx33f07.x
31	24	68.6	190	44	A1254390	A1254390 qv33b05.x
32	24	68.6	177	44	A1254415	A1254415 qv34b11.x
33	24	68.6	262	44	A1254455	A1254455 qv47f11.x
34	24	68.6	176	44	A1254797	A1254797 qv53d08.x
35	24	68.6	395	44	A1254907	A1254907 qv55e02.x
36	24	68.6	190	44	A1255106	A1255106 qv46e04.x
37	24	68.6	176	44	A1266290	A1266290 qp69e05.x
38	24	68.6	146	44	A1271282	A1271282 qp83b03.x
39	24	68.6	160	44	A1271407	A1271407 qv82e11.x
40	24	68.6	257	44	A1284677	A1284677 qv24e08.x
41	24	68.6	61	44	A1284843	A1284843 qv25d10.x
42	24	68.6	61	44	A1305311	A1305311 qv76f07.x
43	24	68.6	156	44	A1305926	A1305926 qv79e04.x
44	24	68.6	126	44	A1306732	A1306732 qv80b09.x
45	24	68.6	109	44	A1306853	A1306853 qv70c06.x

ALIGNMENTS

RESULT	1
AA340759	
LOCUS	AA340759 337 bp mRNA EST 21-APR-1997
DEFINITION	EST46034 fetal kidney of Homo sapiens cDNA 5' end similar to similar to protein phosphatase 2A, beta, 55 kDa, mRNA sequence

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 337)

REFERENCE
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fulder, R.A., Bilt, C.U., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, M.C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fline, D., Fitzgerald, L.M., Fitzhugh, M.M., Fritchman, J.L., Geoghagen, N.S., Glodok, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaro, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y., *Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Fertle, A., Fischer, C., Hastings, G.A., He, W.M., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Welsner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280

TITLE
On Sep 12, 1996 this sequence version replaced gi:1406936.
MEDLINE Other ESTs: EST46035 TRC125647
COMMENT Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13-21.

FEATURES
source
Location/Qualifiers
1..337
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):142407"
/db_xref="taxon:9606"
/clone_lib="Fetal Kidney II"
/dev_stage="fetus"
/note="Organ: Kidney; Vector: pbluescript KS-; Site_1: XhoI; Site_2: EcoRI"
BASE COUNT 63 a 88 c 95 g 88 t 3 others
ORIGIN

Query Match 100.0%; Score 35; DB 32; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTGTGTGTGTGT 35
|||||
Db 4 GACTCGAGTCGACATCGATTTTGTGTGTGT 38

RESULT 2
LOCUS AI295887 678 bp mRNA EST 01-DEC-1998
DEFINITION LP09649, 5prime LP Drosophila melanogaster larval-early pupal P0T2
Drosophila melanogaster CDNA clone LP09649 5prime, mRNA sequence.
ACCESSION AI295887
NID 93945294
VERSION AI295887.1 GI:3945294
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

REFERENCE
AUTHORS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 678)

REFERENCE
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brodstein, P., Lewis, S. and Rubin, C.M.
BDGP/HMT Drosophila EST Project
Unpublished (1997)

TITLE
On Jan 17, 1998 this sequence version replaced gi:2044369.
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
Plate: 96 row E column: 1
High quality sequence stop: 635.

FEATURES
source
Location/Qualifiers
1..678
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="LP09649"
/clone_lib="LP Drosophila melanogaster larval-early pupal P0T2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DHS-alpha"
/note="Organ: whole body; Vector: P0T2; Site_1: EcoRI; Site_2: XhoI. Sized fractionated cDNAs were directly ligated into P0T2. Plasmid CDNA library."

BASE COUNT 204 a 178 c 154 g 142 t
ORIGIN

Query Match 73.1%; Score 25.6; DB 44; Length 678;
Best Local Similarity 87.5%; Pred. No. 2e+02;
Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 TCGAGTCGACATCGATTTTGTGTGTGTGT 35
|||||
Db 621 TCGATCGACCTCGAGTTTGTGTGTGT 590

RESULT 3
LOCUS AI312326 111 bp mRNA EST 14-DEC-1998
DEFINITION ta78e03.x2 NCL CGAP_HSC2 Homo sapiens cdna clone IMAGE:2050204 3', mRNA sequence.
ACCESSION AI312326
NID 94017931
VERSION AI312326.1 GI:4017931
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 111)

TITLE
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
On Feb 10, 1998 this sequence version replaced gi:2340789.
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

ACCESSION C18175
 VERSION 91579777
 C18175.1 GI:1579777
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 435)
 Fujitara, T., Hirano, H., Hishigaki, H., Horie, M., Kawai, A., Kuga, Y., Kyushiki, H., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Suzuki, M., Takai, A., Takeda, S., Matanabe, T., Makawa, H., Nakamura, Y. and Takahashi, E.
 TITLE Otsuka cDNA project
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1392743.
 CONTACT: Tsutomu Fujitara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawasuchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035.
 FEATURES
 SOURCE Location/Qualifiers
 1.435
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GEN-558E05"
 /clone_1lb="Human placenta cDNA (Fujitara)"
 /tissue_type="Placenta"
 BASE COUNT 131 a 88 c 111 g 105 t
 ORIGIN
 Query Match 70.9%; Score 24.8; DB 28; Length 435;
 Best Local Similarity 92.9%; Pred. No. 3.8e+02;
 Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 8 GTGACATCGATTTTATTTTATTTT 35
 ||||| ||||| ||||| ||||| |||||
 Db 274 GTGACATCGATTTTATTTTATTTT 247
 RESULT 9
 AA541358 314 bp mRNA EST 19-AUG-1997
 LOCUS n177d04.s1 NCI_CGAP_Pri2 Homo sapiens cDNA clone IMAGE:982855, mRNA
 DEFINITION sequence.
 ACCESSION AA541358
 NID 92287792
 VERSION AA541358.1 GI:2287792
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 314)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394510.
 CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D., Rodrigo F. Chuang, M.D., Michael R. Emmert-Stuck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 389 Std Error: 0.00
 Seq primer: -40m13 fwd. EP from Amersham
 High quality sequence stop: 273.
 FEATURES
 SOURCE Location/Qualifiers
 1.314
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:982855"
 /clone_1lb="NCI_CGAP_Pri2"
 /sex="male"
 /tissue_type="metastatic prostate bone lesion"
 /lab_host="DH10B"
 /note="Vector: PAMP10; mRNA made from metastatic prostate lesion of the bone, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."
 BASE COUNT 97 a 52 c 65 g 100 t
 ORIGIN
 Query Match 70.3%; Score 24.6; DB 35; Length 314;
 Best Local Similarity 87.1%; Pred. No. 4.8e+02;
 Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 5 CGAGTCGACATCGATTTTATTTTATTTT 35
 ||||| ||||| ||||| ||||| |||||
 Db 313 CGAGTCGACATCGATTTTATTTTATTTT 283
 RESULT 10
 A1726205 466 bp mRNA EST 11-JUN-1999
 LOCUS BNGH5252 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to (Y087741) Delta-tubulin 1 [Hordeum vulgare], mRNA sequence.
 DEFINITION
 ACCESSION A1726205
 NID 95045057
 VERSION A1726205.1 GI:5045057
 KEYWORDS EST.
 SOURCE upland cotton.
 ORGANISM Gossypium hirsutum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
 REFERENCE 1 (bases 1 to 466)
 Blewitt, M., Matz, E.C., Davy, D.F. and Burr, B.
 ESTs from developing cotton fiber
 JOURNAL Unpublished (1999)
 COMMENT On May 18, 1998 this sequence version replaced gi:3137811.
 CONTACT: Ben Burr
 Biology Department
 Brookhaven National Laboratory
 Upton, NY 11973, USA
 Tel: 516-344-3396
 Fax: 516-344-3407
 Email: burrd@bnl.bnl.gov
 Seq primer: T3 primer.
 FEATURES
 SOURCE Location/Qualifiers
 1.466
 /organism="Gossypium hirsutum"
 /cultivar="Acala Maxxa"
 /db_xref="taxon:3635"
 /clone_1lb="Six-day Cotton fiber"
 /tissue_type="immature fiber"
 /dev_stage="Six days post anthesis"
 /lab_host="XL1-Blue"
 /note="Vector: pBluescript II KS+"
 BASE COUNT 133 a 118 c 71 g 144 t
 ORIGIN

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquil,
M.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Kitzman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/dbip/image/image.html

TITLE *The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced g1:1292118.

Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LMU; contact the
IMAGE Consortium (info@image.lmu.gov) for further information.
MGI:371739
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 347.

FEATURES
SOURCE

1. 570
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone_image="606307"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1:
EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGACGTTTTTTTTTTTTTTT 3' "

BASE COUNT 143 a 115 c 137 g 135 t

ORIGIN

Query Match 69.1%; Score 24.2; DB 29; Length 570;
Best Local Similarity 89.7%; Pred. No. 4.9e+02;
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

UY 7 AGTCGACATCGATTTTTTTTTTTTTTTT 35
|| ||||| ||||| ||||| |||||
Db 314 AGTCGACATCGATTTTTTTTTTTTTTTT 342

Search completed: September 13, 1999, 15:45:55
Job time: 4043 sec

Query Match 100.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTAAGCTTAAAGTCACTATCATCC 35
|||||

DB 1 TTTAAGCTTAAAGTCACTATCATCC 35

RESULT 2
BOVPHOS3K/c
LOCUS Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete cds.
DEFINITION
ACCESSION M93252
VERSION M93252.1 GI:163519
KEYWORDS Phosphatidylinositol 3-kinase.
SOURCE Bos taurus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovine.
REFERENCE 1 (bases 1 to 3207)
Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R., Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F., Hsuan, J.J., Courtneidge, S.A., Parker, P.J. and Waterfield, M.D. Phosphatidylinositol 3-kinase: Structure and expression of the 110 kd catalytic subunit
JOURNAL Cell 70, 419-429 (1992)

JOURNAL MEDLINE 92354059

FEATURES
source location/Qualifiers
1..3207
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBR-1"
1..3207
/EC_number="2.7.1.67"
/product="phosphatidylinositol 3-kinase"
/protein_id="AAA30698.1"
/db_xref="PIR:G163520"
/db_xref="GI:163520"

CDS
1
translation="MPRRSSGELMGHILMPRIIVECLIPNGMIVTECLREALTIT
IKHELFKARKYPLHQLDDESSYIFSVTOEAREEDEFETRCDLRLFPPLKVI
EAVNGREKILNREIGFAIGMVECEVDVQEDFRNINLVCKEAVDLNDSPH
SRAMVYPPNVSSEPLPKHLYNKLDKGOIIVYIWIYSPNDKOKYTLKIHDCVPE
OYIAEAIKRTKSMSSBOLKLVLEQGYKYLAVCGDEYFLKPYLSQVYKTRSC
IMLGMPNMLMAKESLSQLEPDCFTMPSTSRISTATPYMNGETSTKSLYINSAL
RIKILCAIVVNNIRIDIKIVRTGIYHGEPLCDNVATQVPCSNPRNENLVNDIY
IPDLPRARLCLISICSVKGRKAKEHCPLANGINLEDYDITLISGKALMLNVPVH
GLEDLNPIGVTSNPNKTEPLEEDFMSVYKPFDMYSIEHANMSVSRGFSY
SHAGLSNRLARDNEIRENDKQLAICTRDPLSETTEDEKDFLMSHRYCAVILPELP
KLISVKNRSDENVQMYCIVDMPRIEPEOMELLDNCPYMPRGFAVRLLEYITP
DDKLSQYLIQVLYKTYEQYLDNLVRLFKALNORIGHEFFHLSKHNKTVSQ
RFGDLLESYCRACGYLKHNLQVAMEKLINLTDLIOEKDETOYOMKFLVEQNR
RPDPMDALOGESLPLNPAHOLGNLRECRIMSARKPLMNMENPDIMSLEIQNR
IIFKNGDLRQMDLTLOIRIMENIMONGDLRLMLPYGCLISGCVGLIEVRSHT
IMQICGGKLGALQFNSHTLHOMLNKNGEYIYDAIDLFTRSCAGVATFLLIGIG
DRHNSNIVKDKGOLFHTDGHFLDKHKKKRYRERVPYVLTQDFLIVISGAGECT
KTRFEROEMCYKAYLAIRQHANLFINLSMGLSGPELQSPDIAYIRKTLALDK
TEQDALETFKQMDANDHGGWTTKMDWIFHTIKOHALN"

BASE COUNT 1028 a 581 c 680 g 918 t

ORIGIN

Query Match 74.3%; Score 26; DB 3; Length 3207;
Best Local Similarity 100.0%; Pred. No. 0.61;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AGGCACTTCAAGTCACTATCATCC 35
|||||

DB 10 AGGCACTTCAAGTCACTATCATCC 35

DB 110 AGGCACTTCAAGTCACTATCATCC 85
|||||

RESULT 3
A37232/c
LOCUS A37232 3498 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9403609.
ACCESSION A37232
VERSION 92294345
KEYWORDS A37232.1 GI:2294345
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3498)
Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D. EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
source location/Qualifiers
1..3498
/organism="unidentified"
/db_xref="taxon:32644"
1..3207
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PIR:E304808"
/db_xref="PIR:G2294346"
/db_xref="GI:2294346"

CDS
1
translation="MPRRSSGELMGHILMPRIIVECLIPNGMIVTECLREALTIT
IKHELFKARKYPLHQLDDESSYIFSVTOEAREEDEFETRCDLRLFPPLKVI
EAVNGREKILNREIGFAIGMVECEVDVQEDFRNINLVCKEAVDLNDSPH
SRAMVYPPNVSSEPLPKHLYNKLDKGOIIVYIWIYSPNDKOKYTLKIHDCVPE
OYIAEAIKRTKSMSSBOLKLVLEQGYKYLAVCGDEYFLKPYLSQVYKTRSC
IMLGMPNMLMAKESLSQLEPDCFTMPSTSRISTATPYMNGETSTKSLYINSAL
RIKILCAIVVNNIRIDIKIVRTGIYHGEPLCDNVATQVPCSNPRNENLVNDIY
IPDLPRARLCLISICSVKGRKAKEHCPLANGINLEDYDITLISGKALMLNVPVH
GLEDLNPIGVTSNPNKTEPLEEDFMSVYKPFDMYSIEHANMSVSRGFSY
SHAGLSNRLARDNEIRENDKQLAICTRDPLSETTEDEKDFLMSHRYCAVILPELP
KLISVKNRSDENVQMYCIVDMPRIEPEOMELLDNCPYMPRGFAVRLLEYITP
DDKLSQYLIQVLYKTYEQYLDNLVRLFKALNORIGHEFFHLSKHNKTVSQ
RFGDLLESYCRACGYLKHNLQVAMEKLINLTDLIOEKDETOYOMKFLVEQNR
RPDPMDALOGESLPLNPAHOLGNLRECRIMSARKPLMNMENPDIMSLEIQNR
IIFKNGDLRQMDLTLOIRIMENIMONGDLRLMLPYGCLISGCVGLIEVRSHT
IMQICGGKLGALQFNSHTLHOMLNKNGEYIYDAIDLFTRSCAGVATFLLIGIG
DRHNSNIVKDKGOLFHTDGHFLDKHKKKRYRERVPYVLTQDFLIVISGAGECT
KTRFEROEMCYKAYLAIRQHANLFINLSMGLSGPELQSPDIAYIRKTLALDK
TEQDALETFKQMDANDHGGWTTKMDWIFHTIKOHALN"

BASE COUNT 1147 a 626 c 722 g 1003 t

ORIGIN

Query Match 74.3%; Score 26; DB 5; Length 3498;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AGGCACTTCAAGTCACTATCATCC 35
|||||

DB 110 AGGCACTTCAAGTCACTATCATCC 85

RESULT 4
HSPH3K/c
LOCUS HSPH3K 3424 bp mRNA PRI 24-AUG-1995
DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
ACCESSION Z29090
VERSION 9472990
KEYWORDS Z29090.1 GI:472990
phosphatidylinositol 3-kinase.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1068)
AUTHORS Volinia, S., Hiles, I., Ormondtroyd, E., Nizetic, D., Antonacci, R., Rocchi, M. and Waterfield, M.D.
TITLE Molecular cloning, cDNA sequence, and chromosomal localization of the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene
JOURNAL Genomics 24 (3), 472-477 (1994)
MEDLINE 95229146
REFERENCE 2 (bases 1 to 3424)
AUTHORS Volinia, S.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig Institute for Cancer Research, 91 Riding House Street, London, W1P 8BT, UK

FEATURES
source Location/Qualifiers
1..3424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="KGla"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CA8333.1"
/db_xref="PID:9472991"
/db_xref="GI:472991"
/db_xref="SWISS-PROT:P42336"
/translation="MPPRSSGELMGHILMPRLIVBCLPENGIVTECLREAVLT IKHLEFKARKYPIHQLDDESSITFVSVDAREDEFEDTRLCEDRLQPEPLKI EPVNNREKILNREIGALIGPVEFPDVKQDEFRNLINVCKEAVDLRLNSHP SRANVYIPPNVSSPELPHKHYLNKLDGQILVIVLVSPNNDOKYTLKINHCVE QVIAEAIKRTKSMILSSEOLKLCVLYOGYILKVCDCDEPLEKYLKSOYKIRSC IMLGPMNLKMAKESLYSOLPMDCFTHMPSYSRISATPYMNGETSKSLAVINSL RIKILCAVYVNLNRDIDKIVRTGIHGGEPLCDNVNTQVPCSNRPMNMLVYDI IPDLPRARCLISCSYKGRKGAKEEHCPPLAMGNINLEDTYDTLVSGMALNMPVH GLEDLNPICVTGSPNKETPCLELEDFDMSVYKFPDMSVIEHANSVREGFSY SHAGLSNRLANDNELRENDKQLAISTRDLSITDQEDFLMSHRYCVTIEILP KLLSVKNSNDEVAQWYCLVKNPPIKPEQAMELDCNTPDPVGRGAVACLEKYL DKLISOYLIOVLYKYEYLDNLVFLKALINORIGHEFFMHLKSEHNNTVSO RFGLLSYCACGMVYKHLNROYEAMEKILNLDILKQEKDETOKYOMKFIYQNR RPDMDALOGFLSPNPAHOLGNLRECRIMSSAKRPLIMNENPDIMSLLFQNN IIFRNGDLRODMLTLITIRLMENIMQOGLDMLPYGCLSIDCVGLIYVNSHT IMQIQCGLGALQFNSHTLHQLKDNKKEIYDAIDLFTRSCAGCVATFLLIG DRHNSNIMVKDQGLFHDGFLDHHKKRGRERYEPVLTQDFLIYISGAOECT KTRREFEOMCYKAYLAIRHNLFLINLSMGLSGMPELQSFDDIAYIKRTALDK TEQALEYFMQNDAHNGWTTKMDWTFHTIKQHALN"

BASE COUNT 1134 a 618 c 709 g 963 t
ORIGIN

Query Match 69.7%; Score 24.4; DB 10; Length 3424;
Best Local Similarity 96.2%; Pred. No. 2.6;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 AGGCATTCTAAAGTCATCATCC 35
|||||
Db 122 AGGCATCTCTAAAGTCATCATCC 97

RESULT 5
LOCUS HSU79143/c 3207 bp mRNA PRI 02-JAN-1997
DEFINITION Human phosphoinositide 3'-hydroxykinase p110-alpha subunit mRNA,
complete cds.
ACCESSION U79143
VERSION U79143.1 GI:1763625
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3207)
AUTHORS Sturdivant, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledder, L.M., Olliff, A. and Heimbrock, D.C.
TITLE Catalytic Activity of the p110-alpha Subunit of Human Phosphoinositide 3'-Hydroxykinase is Required for Signal Transduction
JOURNAL Bioorg. Med. Chem. (1996) In press
REFERENCE 2 (bases 1 to 3207)
AUTHORS Sturdivant, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledder, L.M., Olliff, A. and Heimbrock, D.C.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs, Sunnytown Pike, West Point, PA 19486, USA

FEATURES
source Location/Qualifiers
1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3207
/function="phospholipid kinase"
/product="phosphoinositide 3'-hydroxykinase p110-alpha subunit"
/codon_start=1
/protein_id="A839753.1"
/db_xref="PID:91763626"
/db_xref="GI:1763626"
/translation="MPPRSSGELMGHILMPRLIVBCLPENGIVTECLREAVLT IKHLEFKARKYPIHQLDDESSITFVSVDAREDEFEDTRLCEDRLQPEPLKI EPVNNREKILNREIGALIGPVEFPDVKQDEFRNLINVCKEAVDLRLNSHP SRANVYIPPNVSSPELPHKHYLNKLDGQILVIVLVSPNNDOKYTLKINHCVE QVIAEAIKRTKSMILSSEOLKLCVLYOGYILKVCDCDEPLEKYLKSOYKIRSC IMLGPMNLKMAKESLYSOLPMDCFTHMPSYSRISATPYMNGETSKSLAVINSL RIKILCAVYVNLNRDIDKIVRTGIHGGEPLCDNVNTQVPCSNRPMNMLVYDI IPDLPRARCLISCSYKGRKGAKEEHCPPLAMGNINLEDTYDTLVSGMALNMPVH GLEDLNPICVTGSPNKETPCLELEDFDMSVYKFPDMSVIEHANSVREGFSY SHAGLSNRLANDNELRENDKQLAISTRDLSITDQEDFLMSHRYCVTIEILP KLLSVKNSNDEVAQWYCLVKNPPIKPEQAMELDCNTPDPVGRGAVACLEKYL DKLISOYLIOVLYKYEYLDNLVFLKALINORIGHEFFMHLKSEHNNTVSO RFGLLSYCACGMVYKHLNROYEAMEKILNLDILKQEKDETOKYOMKFIYQNR RPDMDALOGFLSPNPAHOLGNLRECRIMSSAKRPLIMNENPDIMSLLFQNN IIFRNGDLRODMLTLITIRLMENIMQOGLDMLPYGCLSIDCVGLIYVNSHT IMQIQCGLGALQFNSHTLHQLKDNKKEIYDAIDLFTRSCAGCVATFLLIG DRHNSNIMVKDQGLFHDGFLDHHKKRGRERYEPVLTQDFLIYISGAOECT KTRREFEOMCYKAYLAIRHNLFLINLSMGLSGMPELQSFDDIAYIKRTALDK TEQALEYFMQNDAHNGWTTKMDWTFHTIKQHALN"

BASE COUNT 1043 a 586 c 670 g 908 t
ORIGIN

Query Match 69.7%; Score 24.4; DB 10; Length 3207;
Best Local Similarity 96.2%; Pred. No. 2.6;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 AGGCATTCTAAAGTCATCATCC 35
|||||
Db 110 AGGCATCTCTAAAGTCATCATCC 85

RESULT 5
LOCUS MMU03279/c 3207 bp mRNA ROD 25-MAY-1994
DEFINITION Mus musculus Bab/c phosphatidylinositol 3-kinase 110 kDa subunit
mRNA, complete cds.
ACCESSION U03279
VERSION U03279.1 GI:414994
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE 1 (bases 1 to 3207)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eukaryota; Primates; Catarrhini; Homiidae; Homo.

AUTHORS Klippel, A., Escobedo, J.A., Hirano, M. and Williams, L.T.
TITLE The interaction of small domains between the subunits of
JOURNAL Phosphatidylinositol 3-kinase determines enzyme activity
MEDLINE Mol. Cell. Biol. 14, 2675-2685 (1994)
REFERENCE 2 (bases 1 to 3207)
AUTHORS Klippel-Giese, A.
TITLE Direct Subunit
JOURNAL Submitted (05-NOV-1993) Anke Klippel-Giese, CVRI and Dailchi
REFERENCE Research Center, University of California San Francisco, 505
AUTHORS Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
LOCATION/Qualifiers
source 1..3207
FEATURES
source /organism="Mus musculus"
 /strain="Balb/c"
 /db_xref="taxon:10090"
 1..3207
 /standard_name="p110 protein"
 /codon_start=1
 /product="phosphatidylinositol 3-kinase 110 kDa subunit"
 /protein_id="AA1834.1"
 /db_xref="pid:9414995"
 /db_xref="gi:414995"
 /translation="MPPRSSEGLMGTHLMPRIIVCELLPNGMTVTECEAREATLYT
 IKHELFRARKPIQLQDDESTIVTSVTOEAREEPEDETRRLCDLRLPPLIKVI
 EPGVREERKILREIGFAIGMICEFDNKPVEDFRNTLNCKEANDIDNAPH
 SPALVYPPNVSSELPKRIYNKLDKQIIVIVIVSPPNDKOKYTLKINHCVPE
 OVAIAIRKTRSMILSEQLKLVLEOGKYLKVCGDEYFLERKYSQYKIRSC
 IMIGMPNLMMAKESLYQLIDSTFMPYSRISTATPYNGESTSLVYKINSAL
 RLKICATVYVNIIDRIKYRTGTYHNGEPLCNVTPRVCSNPRNEMENLIDY
 IPDLPRALRLCLSTCSVNGRGRKEHCLPANGININLPDYTLVSGKMLAMPVPH
 GLEDLNIGVTSNPKETPCLELFDPMFSSVYKPPKSVIEEHNVSSEAEFYS
 SHTGISNRLARNDRENDKEQLALCTRPDLSETTEQKDLMSHRHCVIPILP
 KLLSVKWNRSDEVAOMYCLVMDPPIKPEQAMELDLQYPPDPAVRSACLELYT
 DDKLSQYLIQVQVLYEQYLDNLVRLKRALNORIGHFFFWHLKSEMNKTVSQ
 RFGLLSEYCSACGMYLKLHROVAMEKILNLTDLKOEKDETOYKMKLVQMR
 OPDEMDALOGELSPINPAHQNLNLECRIMSSAKRRLMNMENDPIMSELTPQNE
 IIFKGGDLROMLTLQIRIMENIWNQGLDRLPYGCLSIGCVGLEIVRNSHT
 IMQICGKGLAQDFNSHTLHOMLKDKNGEITPAIDLFTRSCAGCVATFLLIGT
 DRHNSNINWKDGGDLFDHFDGFLDHKKKFGYKREVRPEVLJODFLIVISGADECT
 KTRFERROEMCYKAYLAIRHANSFTLFSMGLSGMPELQSFDDIAYIRKTLADK
 TEOELAEFTKQMDAHHGCTTKMDYTFHTIKOHALN"

BASE COUNT 958 a 693 c 734 g 822 t

ORIGIN

Query Match 69.7%; Score 24.4; DB 12; Length 3207;
Best Local Similarity 96.2%; Pred. No. 2.6;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 10 AGGCATCTTAAGTCACTATCATCC 35
 |||||
 |||||

Db 110 AGGCATCTTAAGTCACTATCATCC 85

RESULT 7
AF001076/c 3452 bp mRNA VRT 08-JUL-1997
LOCUS Gallus gallus phosphoinositide 3-kinase catalytic subunit mRNA,
DEFINITION complete cds.
ACCESSION AF001076
NID 92245505
VERSION AF001076.1 GI:2245505
KEYWORDS
SOURCE chicken.
ORGANISM Gallus gallus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
AUTHORS Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 1 (bases 1 to 3452)
 Chang, H.W., Aoki, M., Furman, D., Auger, K.R., Bellacosa, A.,
 Tschilis, P.N., Cantley, L.C., Roberts, T.M. and Vogt, P.K.
TITLE Transformation of chicken cells by the gene encoding the catalytic
 subunit of PI 3-kinase

JOURNAL Science 276 (5320), 1848-1850 (1997)
MEDLINE 97354438
REFERENCE 2 (bases 1 to 3452)
AUTHORS Chang, H.W., Aoki, M. and Vogt, P.K.
TITLE Direct Subunit
JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
REFERENCE Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
AUTHORS BCC-239, La Jolla, CA 92037, USA
LOCATION/Qualifiers
source 1..3452
FEATURES
source /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /tissue_type="brain"
 1..150
 151..3354
 /gene="c-p3k"
 /note="p3k proto-oncogene"
 151..3357
 /note="c-p3k"
 /codon_start=1
 /product="phosphoinositide 3-kinase catalytic subunit"
 /protein_id="AAB62534.1"
 /db_xref="pid:92245506"
 /db_xref="gi:2245506"
 /translation="MPPRSSEGLMGTHLMPRIIVCELLPNGMTVTECEAREATLYT
 IKHELFRARKPIQLQDDESTIVTSVTOEAREEPEDETRRLCDLRLPPLIKVI
 EPGVREERKILREIGFAIGMICEFDNKPVEDFRNTLNCKEANDIDNAPH
 SPALVYPPNVSSELPKRIYNKLDKQIIVIVIVSPPNDKOKYTLKINHCVPE
 OVAIAIRKTRSMILSEQLKLVLEOGKYLKVCGDEYFLERKYSQYKIRSC
 IMIGMPNLMMAKESLYQLIDSTFMPYSRISTATPYNGESTSLVYKINSAL
 RLKICATVYVNIIDRIKYRTGTYHNGEPLCNVTPRVCSNPRNEMENLIDY
 IPDLPRALRLCLSTCSVNGRGRKEHCLPANGININLPDYTLVSGKMLAMPVPH
 GLEDLNIGVTSNPKETPCLELFDPMFSSVYKPPKSVIEEHNVSSEAEFYS
 SHTGISNRLARNDRENDKEQLALCTRPDLSETTEQKDLMSHRHCVIPILP
 KLLSVKWNRSDEVAOMYCLVMDPPIKPEQAMELDLQYPPDPAVRSACLELYT
 DDKLSQYLIQVQVLYEQYLDNLVRLKRALNORIGHFFFWHLKSEMNKTVSQ
 RFGLLSEYCSACGMYLKLHROVAMEKILNLTDLKOEKDETOYKMKLVQMR
 OPDEMDALOGELSPINPAHQNLNLECRIMSSAKRRLMNMENDPIMSELTPQNE
 IIFKGGDLROMLTLQIRIMENIWNQGLDRLPYGCLSIGCVGLEIVRNSHT
 IMQICGKGLAQDFNSHTLHOMLKDKNGEITPAIDLFTRSCAGCVATFLLIGT
 DRHNSNINWKDGGDLFDHFDGFLDHKKKFGYKREVRPEVLJODFLIVISGADECT
 KTRFERROEMCYKAYLAIRHANSFTLFSMGLSGMPELQSFDDIAYIRKTLADK
 TEOELAEFTKQMDAHHGCTTKMDYTFHTIKOHALN"

BASE COUNT 3'UTR
BASE COUNT 1073 a 672 c 783 g 924 t

ORIGIN

Query Match 65.1%; Score 22.8; DB 4; Length 3452;
Best Local Similarity 92.3%; Pred. No. 11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 10 AGGCATCTTAAGTCACTATCATCC 35
 |||||
 |||||

Db 260 AGGCATCTTAAGTCACTATCATCC 235

RESULT 8
AF001075/c 3389 bp mRNA VRT 08-JUL-1997
LOCUS Avian sarcoma virus 16 gag-v-phosphoinositide 3-kinase catalytic
DEFINITION subunit fusion protein (gag-v-p3k) and retroviral env mRNAs,
 partial cds.
ACCESSION AF001075
NID 92245502
VERSION AF001075.1 GI:2245502
KEYWORDS
SOURCE Avian sarcoma virus 16.
ORGANISM Avian sarcoma virus 16
REFERENCE Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
 1 (bases 1 to 3389)
 Chang, H.W., Aoki, M., Furman, D., Auger, K.R., Bellacosa, A.,
 Tschilis, P.N., Cantley, L.C., Roberts, T.M. and Vogt, P.K.

TITLE Transformation of chicken cells by the gene encoding the catalytic subunit of p13-kinase

JOURNAL Science 276 (5320), 1848-1850 (1997)

MEDLINE 97334438

REFERENCE 2 (bases 1 to 3389)

AUTHORS Chang, H.W., Aoki, M. and Vogt, P.K.

TITLE Direct Submission

JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop BCC-239, La Jolla, CA 92037, USA

FEATURES Location/Qualifiers

source

1. .3389

/organism="Avian sarcoma virus 16"

/db_xref="taxon:60629"

<1. .3252

/gene="gag-v-p3k"

/note="fusion gene of gag and v-p3k"

1. .78

/gene="gag-v-p3k"

/note="derived from gag gene"

<1. .3252

/gene="gag-v-p3k"

/codon_start=1

/product="gag-v-phosphoinositide 3-kinase catalytic subunit fusion protein"

/protein_id="AAB62532.1"

/db_xref="PDB:92245503"

/db_xref="GI:2245503"

/translation="QQLIRAPSTLTTPGELIKYVLDKQKASHLMPRLVCELPNG MIVLECEKRETLITIKHELEKARKYPLYLODDESYIPVSTOAEEREFEDTR RCDLRLOPRLKYIPVNGEPEKILRKISFAIGMPCEDMDKDEVDQFRNITL VCKEAVLDKRNAMHSHALVCPNVSSEPELPHNINKDKGITVAVINVTSPND KOKTILINDCYEVOYIAEIRKTSMLSSQIKLCVLEYGKYLKVGODEYL LEKPLSQVYTRSCIMGRPNMLMAKESITQPLDFTMPSYSKRISTATPYM GEATKSLYNTINSALIRIKATVANNINIDIKIVTGTGEGEPLCDVNTORVP CSNRMNEMSYDMYIPDLPAPRALCLSGVGRKGAKEEGFLAMNINMEYTOT LVSGKALNTMAVPHGLEDLNPIGATGNSNKRKTCLELEFDMFSNPKRPMYIE EHNMTISRELGFTYAGUSNRIRAKNRELSEKQRAICTDPSLSEITQKDEL WSHRYCVKTEPELIPKLLSVKNSRDEVAQMYLVDMPIKEQAMELLDCYPPD MRAFAVRCLEKYLTDKLSQYLQVAVLEEDYLDNLPFLKALNLTORGHF FHLKSEMHNTVQORGLLESYCRAGMGLKLSROVEAMEKILNLTOKERD ETQVOKKFLVEQMRDEMDALOGFISPLPAHOLGLREECRIMSSARPMILNK ENDMISELLPNNELLFKNGDILROMLTIQIRIMNTWONGDLRLMPTCLST GDCVGLLEVRSSTIMQIQCKGKGLKALQFNSTLRQWLNKNGEATDAIDLETR SCAGYCVATFLLIGDRHNSNINWKGQLEFHDGFLDKKKKFGKRRVPEVLT QDLDIVISKGAOECTKREFEFQEMCHKAYLAIRHANLEINLFSMMLGSGMPELOS FDIAYIRKTLALDKTEQEALEYFMKQNDHGGWTKMWMIFHTIKOHALN"

79. .84

/gene="gag-v-p3k"

/note="sequence of unknown origin"

85. .3252

/gene="gag-v-p3k"

/note="derived from c-p3k gene"

3250. .3370

3371. .>3389

/codon_start=1

/product="retroviral env"

/protein_id="AAB62533.1"

/db_xref="PDB:92245504"

/db_xref="GI:2245504"

/translation="MARNS"

BASE COUNT 1071 a 655 c 751 g 912 t

ORIGIN

Query Match 65.1%; Score 22.8; DB 17; Length 3389;
Best Local Similarity 92.3%; Pred. No. 11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 AGGCAATCTAAAGTCACTATCATCC 35
||||||| ||||||| ||||||| |||
Db 155 AGGCATCTAGAGTCACTATCATTC 130

RESULT 9

CETO4A8 39482 bp DNA INV 23-NOV-1998

LOCUS Caenorhabditis elegans cosmid T04A8, complete sequence.

DEFINITION 235653

ACCESSION 9527419

NID 9527419

VERSION 235653.1 GI:527419

KEYWORDS HTG; 1,4-Alpha-glucan branching enzyme; 50S ribosomal protein L16; Amidophosphoribosyltransferase; C-type lectin; calpain thiol protease; DNA; emb-5 gene; glucose-6-phosphate 1-dehydrogenase; Glutamine phosphoribosylpyrophosphatase; amidotransferase; neurofilament triplet M protein; Non-histone chromosomal protein HMG-1.

SOURCE

ORGANISM Caenorhabditis elegans.

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Pelodertinae; Caenorhabditis.

AUTHORS 1 (bases 1 to 39482)

JOURNAL Palmer, S.

REFERENCE Direct Submission

AUTHORS Submitted (04-AUG-1994) Louis, MO 63110, USA. E-mail: jesus@anger.ac.uk or rwenematode.wustl.edu

REFERENCE 2 (bases 1 to 39482)

AUTHORS Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berts, M., Bonfield, J., Burton, J., Connell, M., Copsey, J., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Riken, L., Roopra, A., Saunders, D., Showkneen, R., Smalton, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkison, Sproat, J. and Woldman, P.

TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans 368 (6466), 32-38 (1994)

JOURNAL Nature 368 (6466), 32-38 (1994)

MEDLINE 94150718

COMMENT Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see: -
http://webc.sanger.ac.uk/cgi-bin/display?db=wormacc&class=sequence&object=T04A8
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone T04A8. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone T04A8 is at 1 in this sequence. The true right end of clone T04A8 is at 4879 in sequence Z35719.
The true right end of clone T23F11 is at 5962 in this sequence. The start of this sequence (1. .96) overlaps with the end of sequence Z46343.
The end of this sequence (39470. .39482) overlaps with the start of sequence Z35719.
Location/Qualifiers

1. .39482

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="III"

/clone="T04A8"

```

gene      285..2182
/ gene="arg-10"
CDS       join(283..515,1092..1310,1365..1508,1567..1749,1985..2182)
/ note="similar to 7tm receptor"
/ codon_start=1
/ protein_id="CAA84725.1"
/ db_xref="PID:e1348903"
/ db_xref="PID:g3879340"
/ db_xref="GI:3879340"
/ db_xref="SWISS-PROT:P46568"
/ translation="MNSSRVPPANFSEYEDPLPEECNEDPNVVLSTAMYGMSYLIY
GAVLNAVITYVEHGNSYRDNFPMYCYADATVGYINTAEVIFGRIFETIPIPIPIA
SPYFTPSLITKMYAALHSYSGFKTSQIFMSFRNMTVITLAKELIKWQILKLEPV
LITFTLPLGVNKLILSRVYINPNNGAFSVNKKDFPANNISILHLEHETLFCVAVI
FVATILGLTLMKORIKSAERSLITVTMAVQVYFASIQIYFVEFAAYTERKISVLI
LOIVSEVEDSLVVFSEPIALIVMSROLRKIDIFMLKDETQISMVPSNEL"
gene      2532..3697
/ gene="arg-11"
CDS       join(2532..2658,2878..3338,3506..3697)
/ gene="arg-11"
/ note="See T23F11.5"
/ codon_start=1
/ protein_id="CAA84735.1"
/ db_xref="PID:e1348912"
/ db_xref="PID:g3879350"
/ db_xref="GI:3879350"
/ db_xref="SWISS-PROT:P46569"
/ translation="MSEFRPNISNDITIIIFECNSNYDIVEYKMFLOJAVLIFGI
LNLILLYTLFKNSEIYVASSFEFLYSDCFVSMIFLDIGRTLYVETPTPIAP
MEYEPIDGRIMVILVHRSACKSLIQLLYVNMSCVYPIRYGKMRPKYIILV
VFVPEPIDGNMLISVYMQPREGCIYMEYIKVMASSRQLIIFTIALLFTVYCT
SVIFTVLVPRKLRNVERTLSLGRYISMSFTIIVFQVN"
gene      complement(3293..7300)
CDS       / gene="T04A8.3"
/ complement(join(5293..5696,5899..6060,6234..6659,
6706..6965,7019..7072,7227..7300))
/ note="similar to C-type lectin"
/ codon_start=1
/ protein_id="CAA84726.1"
/ db_xref="PID:e1348903"
/ db_xref="PID:g3879341"
/ db_xref="GI:3879341"
/ db_xref="SPTREMBL:Q22136"
/ translation="MFDVTIVRMLESIDIFKMRNSTRMKVTNARSSFLNRTRN
PLSTCEDVPPVPHINNAESEYKTCCEATLICSQVQMGRCYKITKMTEDKA
BEHCKNOQDHTSTIAFMHREALPFMNDEYFVSRLIMDASKVITINDLYDEGVNL
LALDGYKYNLPVAVIARVPKDELTANVLCIEYTPPKTSNNTLLRRYGEIYVPLVTS
SVYMTTSSRIRINADPLADNYCTELMKPVYRGEAQSALPTQEFYKKTJCTNFRF
YKRYTGOLMCKHYTHDMSNNDNAERKNDGAHLSSFTYEELKLDLDEMTLEYVPN
DNNAIVWLGAKRRECGDLSKNFTGYSKDIDHPARSVFEVMQVQVONPFIYGDG
EDYMAKEYEPNSTDSERCLVQMSGMSVYGDNPRNMQINDIYCNTEYFKLCGEA
TFYAVKKNL"
gene      complement(7871..9714)
CDS       / gene="T04A8.4"
/ complement(join(7871..8071,8723..8819,8916..9181,
9228..9359,9489..9585,9638..9714))
/ gene="T04A8.4"
/ codon_start=1
/ protein_id="CAA84722.1"
/ db_xref="PID:e1348898"
/ db_xref="PID:g3879337"
/ db_xref="GI:3879337"
/ db_xref="SPTREMBL:Q22133"
/ translation="MPISASDTIRKFNVAANOANEVLKKNPYSDTYKIOAFDKNY
RPPGSKTEARIGIKAGVHCRIELFLCEIDSNADGEPHKYKFKGLNPISEYSDK
LVGMILARKYGLVFESEMLYORODDEKITTFMMSLDEIRSLASGDPANCIOIR
NSEPVATVPVSKELPRPPIRNSDSVTSKAFKAPPPNDISLVQGPKNPVPKID
TSKAPILIRRRRLKIFMGLCLCYTNQVDRDMSFQDMCEMGGEQTYRANGKYQV
IYDLSIDIMLEN"
gene      11864..14597
/ gene="T04A8.5"

```

```

CDS       join(11864..11952,12576..12981,13028..13288,13524..13833,
13882..13997,14293..14457,14502..14597)
/ gene="T04A8.5"
/ note="Amidophosphoribosyltransferase (glutamine
phosphoribosylpyrophosphate amidotransferase)"
/ codon_start=1
/ protein_id="CAA84723.1"
/ db_xref="PID:e1348900"
/ db_xref="PID:g3879338"
/ db_xref="GI:3879338"
/ db_xref="SPTREMBL:Q22134"
/ translation="MCGIFGIVPNSHLNVAANGIAALOHRESTGIVGSDGIT
RDHTITIGHLVADYITEDNISRNNGSITIIHNRKSTACKKSGLNCVQPVYITM
MGVIAIANGELVDAKQKREVLHEGVSLSDTDSLEIAQMIARAIALNVRCKYQEM
GDITRELVNLSAMNSYSLVMTFEDRLYLARDPGRNPLCVGYVSKNCPDEAFAS
SESCAFPNALADEVRGEIYELSTGIRKVMQKPTPLAMCFEYVYPARNDSEI
EGQOVYREEGQTMALDEDLVDIYNVDSSLSAIGVASOGIYEVLHNSY
VGRSPTEPNDMROMNATKMGVYKKYHGORIYVNDSTYRCGTMRTYVRLMDACA
KEVHRIASPVYKFCFPGINIPPTKELIAAEKITPEICQFVGADSVRYLSVDGLVSS
VQKIERANFSPGHCCTACTGKTPVALDL"
/ complement(14747..14836,14881..15396,15445..15558,
15604..15807)
/ gene="T04A8.6"
/ complement(join(14747..14836,14881..15396,15445..15558,
15604..15807))
/ note="similar to ribonucleoprotein; cDNA EST yk222a11.3
comes from this gene; cDNA EST yk432f10.3 comes from this
gene; cDNA EST yk432f10.5 comes from this gene; cDNA EST
yk432f10.5 comes from this gene; cDNA EST yk497a8.3 comes
from this gene; cDNA EST yk497a8.5 comes from this gene"
/ codon_start=1
/ protein_id="CAA84724.1"
/ db_xref="PID:e1348901"
/ db_xref="PID:g3879339"
/ db_xref="GI:3879339"
/ db_xref="SPTREMBL:Q22135"
/ translation="MVAONPSSISGGRKOPKQDKYQKQKTIYVYKIRIPFGE
EKELGYRQGVNLRIVANRRRTGHNKGAIVGFDKAAVLAELTAASMNGLTFEOR
LGCTVMKPELIPKARHGPVLVMPSPSYGIKKPTIANNITGNDVYNAKRVQAN
KTLRKLQMGNGINDSVSGTPTPLKVIDEVOFISDVPKSTPIIVKAAATPKAT
PKAVTPKQATPKVSTPYTKVYIKTPTEKEPQPTPKTAKKAAAAATPAKSTL
SNTILKSVASARBAVAEKKTLRSRKKKS"
gene      15523..17601
CDS       / gene="T04A8.8"
/ join(15523..16784,16831..17000,17047..17601)
/ gene="T04A8.8"
/ note="cDNA EST yk339d3.3 comes from this gene; cDNA EST
yk339d3.5 comes from this gene"
/ codon_start=1
/ protein_id="CAA84734.1"
/ db_xref="PID:e1348911"
/ db_xref="PID:g3879349"
/ db_xref="GI:3879349"
/ db_xref="SPTREMBL:Q22144"
/ translation="MDSDDNAYSDYIGNRRAALQKKRPVRYCESDDSDDEFTSGP
SPARKRTREIOLFPDSDDECEVEKEVETISQKRRHLEDVDEVARISIKREYVIA
PVERNNVTSIGLKICEGMDENAKSGAOPFEVEDYVDDSDDESSIAENFPTVILDC

```

Query Match 65.1%; Score 22.8; DB 36; Length 39482;

Best Local Similarity 79.4%; Pred. No. 6.4;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TTTAAGCTTAGGCATCTAAAGTCATCATCC 34

DB 4747 TTTCAAGATTGGCATTCTAAAGTCATCATTAACC 4714

RESUTL 10
AF076243
LOCUS AF076243 99384 bp DNA PLN 11-MAY-1999
DEFINITION Arabidopsis thaliana BAC T26N6 from chromosome IV at 19.3 CM,
complete sequence.
ACCESSION AF076243

```

XID      93309276
VERSION  AF076243.1  GI:3309276
KEYWORDS
SOURCE    HG.
ORGANISM  Arabidopsis thaliana
           thale cress.
           Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
           eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
           Arabidopsis.
REFERENCE 1 (bases 1 to 99384)
AUTHORS   Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.
TITLE      Genomic sequence of Arabidopsis thaliana BAC T26N6, chromosome IV,
           19.3 cM
JOURNAL    Unpublished
REFERENCE 2 (bases 1 to 99384)
AUTHORS   Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.
TITLE      Direct Submission
JOURNAL    Submitted (02-JUL-1998) Applied Biosystems, Division of Perkin
           Elmer, 850 Lincoln Center Drive, Foster City, CA 94004, USA
           3 (bases 1 to 99384)
REFERENCE 3 (bases 1 to 99384)
AUTHORS   Parnell,L.D.
TITLE      Direct Submission
JOURNAL    Submitted (11-MAY-1999) Lita Annenberg Hazen Genome Sequencing
           Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
           Harbor, NY 11724, USA
REMARK     Arabidopsis thaliana BAC T26N6 from chromosome IV at 19.3 cM
           BAC T26N6 was sequenced as part of the Arabidopsis genome
           sequencing effort of the Cold Spring Harbor Consortium. For
           additional information, please see http://www.cshl.org/arabweb.
           T26N6 is known to carry the m1306 marker. Fingerprint data indicate
           that T26N6 overlaps with T19B17 and F4H6 of YAC C1C3F1.
FEATURES
   source
       1..99384
       /organism="Arabidopsis thaliana"
       /cultivar="Columbia"
       /db_xref="taxon:3702"
       /chromosome="IV"
       /map="19.3 cM"
       /clone="BAC T26N6"
       /note="hybridizing YAC is C1C8B1"
   misc_feature
       1..9514
       /note="Overlap with T19B17, GenBank accession number
           AF069441, from position 1 to 9513; there is a 1-bp
           discrepancy between T26N6 and T19B17"
       4268..4587
       /note="function-unclassified; similar to K23U20, GenBank
           accession number AB016874; similar to T19K24, GenBank
           accession number AC002342"
       /rpt_type="dispersed"
       /rpt_type-dispersed
       5477..6145
       /gene="T26N6.1"
       /note="similar to histone H2A; the potential protein has a
           premature COOH terminus, an extremely low similarity to
           other H2A histones at the amino terminus, and similarity
           to histones on both strands of T26N6"
       /pseudo
       /evidence-not_experimental
   repeat_region
       10402..10406
       /note="function-putative-target duplication site; cognate
           A of target duplication site flanking retrotransposon from
           position 10402 to 15747; cognate B resides from position
           15743 to 15747"
       /rpt_type="flanking"
       10402..15747
       /note="function-putative-retrotransposon"
       /rpt_type="dispersed"
       10407..10796
       /note="cognate A of LTR flanking retrotransposon from
           position 10402 to 15747; cognate B resides from position
           15353 to 15742"
       /gene="T26N6.2"
       /note="encodes putative polypeptide; gene model last
           edited on 4 May 99"
       /evidence-not_experimental
       join(10943..12205,12445..12933,13131..13410,13487..14508)
       /gene="T26N6.2"
       /note="functional catalog ID=90"
       /codon_start=1
       /evidence-not_experimental
       /product="putative polypeptide"
       /protein_id="AAD29768.1"
       /db_xref="PID:94773895"
       /db_xref="GI:4773895"
       /translation="MKNPKAITSVYLKGVYLLMRTKYVLCGRGLMSEYERKHV
           FKATKEDGDSGKRESEAELEKAKFQEDONVALQHSLEIASVCEYAR
           LMETLENYGVNLTVEFEYKAINNLQVDFLEFKFGRFLMELMLRSTVD
           PALINERKOEKQVGLLLTLNPAENDIKHLRLADKLPSLENVCSQVKEGSLGS
           GKGLIANKKQVIGKGRKVVNCDCHGKMRKQVKNKFLPAKAEKVIQSEVAS
           DQGEVVRSDLESIRISLASIKESGSEFLYEPKMKLKEGSEFTSEPSFTLIDSG
           ASHMINNPSILINDIKRALGVYANDKQVYKVEIGELNFDKSKALYMSFTSNLL
           SVKRAINDLNGTYLFGPNSVHFQDKGRSLTILENCDLVHSDVWTSPOKSRNKT
           FVTEIDERSKTYWITLLPSKDCVMDAEINQNYVTNFMAKIKVLRDNGEYSHK
           KEYLAKHGIIHQISCPYTPQONGVAERKNRILMEVSRMHTVTKKFWMDAVMTAC
           YLINRTPKVLQDVSTFEFEKGYEKQDNLSADSPETIDRATSLQFLDHLGVT
           PSSRERKTRDLREPTITIDQENQEPASNLQDQGEINGIQIHDMQTONQDGEVIG
           LREKSSRLTYNNKAVARPIQAVCSLLLPQDHQAFICKTANFVLQYTEKSEERI
           NAVADETAMTRNHTWDEDLPGKRAVSSWVFTITLSGELTERKARLVAGFTQ
           TYGRDYETETAPVAKLHTVRVLSLATNLSDWLMQMKVKNKFLQGELEEVYTPPG
           LEDSIAGKIVLRKRAIKYIGLKOSPRAWHNLSTYLGKGEKFSADHLYFLQSDGI
           IVALIYVNDITISDNGKEGIDHTRFLKSTFEDIDGLKYLFLIYCORSPEGFLSQ
           RKYTLDLNENGKIGSKPAKPIVDGYTKRTGGRRRMHLHR"
       15353..15742
       /note="cognate B of LTR flanking retrotransposon from
           position 10402 to 15747; cognate A resides from position
           10407 to 10796"
       15743..15747
       /note="function-putative-target duplication site; cognate
           B of target duplication site flanking retrotransposon from
           position 10402 to 15747; cognate A resides from position
           10402 to 10406"
       /rpt_type="flanking"
       17025..17085
       /note="function-mitochondrial repeat; similar to T1V1
           element, GenBank accession number X91202"
       /rpt_type="dispersed"
       /rpt_type-dispersed
       complement(19590..19593)
       /note="function-putative-target duplication site; cognate
           A of target duplication site flanking retrotransposon from
           position 19590 to 25423; cognate B resides from position
           25420 to 25423"
       /rpt_type="flanking"
       complement(19590..25423)
       /note="function-putative-retrotransposon"
       /rpt_type="dispersed"
       complement(19594..20140)
       /note="cognate A of LTR flanking retrotransposon from
           position 19590 to 25423; cognate B resides from position
           24873 to 25419"
       complement(21779..24805)
       /gene="T26N6.3"
       /note="encodes putative transposon protein; gene model
           last edited on 1 May 99"
       /evidence-not_experimental
       complement(21779..24805)
       /gene="T26N6.3"
       /note="functional catalog ID=90"
       /codon_start=1
       /evidence-not_experimental
       /product="putative transposon protein"
       /protein_id="AAD29754.1"
       /db_xref="PID:94773881"
       /db_xref="GI:4773881"
       /translation="MDYKPEFVAGKTIULEKGVYGMKVKRALLIGLQKEAMLIAT
           IGRKAPVTKGDEGVDLTKIQNDADAERKAKNSRALSIFNRYNQNQRFQNCES
           AKEMWDKLAKVBEGLSSVKSRSRIDLMSQFENLSMEETENIEEESGKISALIASAHNL

```

```

GKKYKDKLTVKLLKLPSPFESKRTAMGTSLSLDTSDIDEVYVGMQAVELETSGKG
GSKGLATALASAKNEIOELKDTMSMAADSSAMRVRKGGRNQGDRTDRDRSSK
REIOCHECOGHRIKAECPSLRKDKLSECKGCHTFEDGSGSKDPKSSSE
SUNDSDESDYIKGFVGIIEKDESSDSDEGDESDNADSDIDIKDVNIIEF
RLYDSWMLSKREKVMLEBKLYOELTKRGELAAONKSELJOKSVAEKRE
LQELSDTRKKNHMTNSGKGLDLSILAAKRVGNSNGIDYNACSGTKNNPVSAA
PKSCQGFPSNNDVAPARYONHHDYOSRRYTGTECYTCHGHGIIQXCRYARL
NLKROGKLPYOTGRTSKMVRREDYICVAATYSIEGILKPYEPDGSARHVSQS
NLENTYSVESKVTGEGGDKRKIKGGLDLPKAPQLTWYVPEGLTANLISVQLCD
EGILVSENVKCATNEKNQNTLGVTRNOCNMEPEKCELAEKEDVYVNIIEF
MARSMSEIVSKMVRGVOELKIEKIVDAYOQGIROYOHRVYGVVERKQTOE
MARAMTHGVPEKFTWTEASTACYNIHYVYIGCTPKKIYAAFYKTIETERTSM
GELKYRGLQINDINGIAISQSTYQNLVKRPDMGSSNPVEIPHSTIYLCSSITIL
MKHMGGLDGMSSDPLVACDNESAIALSKNPVQSTIKHIALRHFFVELVEEKOI
TVEHVPEIOLDIFKPLDLMFVNLKSLSGEV"
complement(24873..25419)
/feature="cognate B of LTR flanking retrotransposon from
position 19590 to 25423; cognate A resides from position
19594 to 20140."
repeat_#region
complement(25420..25423)
/feature="function-putative_target_duplication_site; cognate
B of target duplication site flanking retrotransposon from
position 19590 to 25423; cognate A resides from position
19590 to 19593."
LTR
/feature="flanking
27302
/feature="function-putative_target_duplication_site; cognate
A of target duplication site flanking retrotransposon from
position 27298 to 36385; cognate B resides from position
36381 to 36385."
repeat_region
27302
/feature="flanking
27302
/feature="function-putative_target_duplication_site; cognate
A of target duplication site flanking retrotransposon from
position 27298 to 36385; cognate B resides from position
36381 to 36385."
LTR
/feature="flanking
27302
/feature="function-putative_target_duplication_site; cognate
A of target duplication site flanking retrotransposon from
position 27298 to 36385; cognate B resides from position
36381 to 36385."
gene
/feature="cognate A of LTR flanking retrotransposon from
position 27298 to 36385; cognate B resides from position
35910 to 36380."
complement(30204..32291)
/feature="126N6.4"
/feature="encodes putative transposon protein; gene model
last edited on 1 May 99"
/feature="not_experimental
evidence-not_experimental
complement(30204..30467,30979..31200,31283..31588,
31687..31712,32057..32291))
/feature="126N6.4"
/feature="functional catalog ID=90"
/feature="codon_start=1
evidence-not_experimental
/product="putative transposon protein"
CDS
/feature="cognate A of LTR flanking retrotransposon from
position 27298 to 36385; cognate B resides from position
35910 to 36380."
complement(30204..32291)
/feature="126N6.4"
/feature="encodes putative transposon protein; gene model
last edited on 1 May 99"
/feature="not_experimental
evidence-not_experimental
complement(30204..30467,30979..31200,31283..31588,
31687..31712,32057..32291))
/feature="126N6.4"
/feature="functional catalog ID=90"
/feature="codon_start=1
evidence-not_experimental
/product="putative transposon protein"
Query Match 61.1%; Score 21.4; DB 8; Length 99384;
Best Local Similarity 80.6%; Pred. No. 19;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
2: 1 TTACCTTACGATCTTAAGTCACTATCA 31
DB 10547 TTTCATCTTAGCCATGTCATCACTATCA 10577

```

```

SOURCE
ORGANISM
Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Archiascomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
REFERENCE
1 (bases 1 to 38078)
McDougal, R.
Unpublished
2 (bases 1 to 38078)
Barrell, B.G., Raeburn, M.A. and Wood, V.
Direct Submission
TITLE
Submitted (17-SEP-1997) Schizosaccharomycetes pombe chromosome I
sequencing project, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk
NOTES
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S.pombe/)
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the GeneFINDER program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites. CDS are numbered using the following
system eg SPAC5H10.01c. SP (S. pombe), A (chromosome 1), c5H10
(cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter, because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid c9E9 is
overlapped at the 5' end by cosmid c17C9.
FEATURES
source
location/Qualifiers
1..38078
organism="Schizosaccharomycetes pombe"
strain="972h-"
db_xref="taxon:4896"
chromosome="I"
clone="cosmid c9E9"
map="I"
118..458
gene="SPAC9E9.01"
join(118..154,196..458)
gene="SPAC9E9.01"
note="SPAC9E9.01, len:94"
codon_start=1
product="very hypothetical protein"
protein_id="CAB16401.1"
db_xref="PID:e1370316"
db_xref="PID:g4154093"
db_xref="GI:4154093"
db_xref="SPTREMBL:O14287"
translation="MADGELISAEISLVTGTVHTPIYVPLVGLGSKRKRKTSQSWVGS
VTKLEAGFVSRCEAANGEMWEDNEDRCRLYQIYRSKIKLKEFARIYVLSRVNM"
155..160
gene="SPAC9E9.01"
note="gtacgc, splice donor sequence"
180..195
gene="SPAC9E9.01"
note="ctaataatagctcaag, splice branch and acceptor"
complement(877..1170)
note="Old SPAC9E9.02"
5353..7629
gene="SPAC9E9.03"
5353..7629
gene="SPAC9E9.03"
note="SPAC9E9.03, len:758, SIMILARITY: Saccharomycetes
cerevisiae, LEU2 YEAST, 3-isopropylmalate dehydratase,
(779 aa), fasta scores: opt: 2607, E(): 0, (63.0% identity
in 774 aa)"

```

[illegible]

Query Match Best Local Similarity 76.5%; Pred. No. 28; Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TTTAGCTTAGGCATCTTAAGTCACATCAATCC 34
||||||| | || |||||
Db 10971 TTTAGCTTAACTTTAGCAGAAGACTTTCAATCC 10938

misc_feature complement(14706..14729)

misc_feature 60.6%; Score 21.2; DB 7; Length 38078;
Best Local Similarity 76.5%; Pred. No. 28;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

gene /product-"theonine synthase"
/protein_id="CAB16415.1"
/db_xref="PID:e1198281"
/db_xref="PID:g2656015"
/db_xref="GI:2656015"
/db_xref="SWISS-PROT:P042598"

misc_feature complement(12305..12349)
/gene="thc"
/note="P500165 Serine/threonine dehydratases
pyridoxal-phosphate attachment site"
complement(12599..12619)
/gene="thrc"
/note="ctaatgaccttactatag, splice branch and acceptor"
complement(12651..12653)
/gene="thrc"
/note="gatgt, splice donor sequence"
complement(14172..14774)
/gene="yp12"
complement(14172..14774)
/gene="yp12"
/note="SPAC969.07c, len:200"
/codon_start=1
/product="yp1-related protein 2."
/protein_id="CAB16405.1"
/db_xref="PID:e349391"
/db_xref="PID:g2414662"
/db_xref="GI:2414662"
/db_xref="SWISS-PROT:P17609"
/translat="MSTRSYDYLKLLIGDSGVKSCLLRFSEDSFTSFITIGI
DFKRTEIDGRIKIIMPTADGERRTITTAAYRAMGLIEDVTDKRSFNVNTR
MFSEVPEHASENVKKILIGNKCEDEDRVSFEQGALDELGVKFLEASAKIVAND
EAFETLARETKRQIKIDENERSNNANVDLDNDTRYARCC"

RESULT 12
AC005350 AC005350 DNA PRI 31-JUL-1998
LOCUS Homo sapiens chromosome 5, BAC clone 261j17 (DBNL H190), complete
DEFINITION sequence.
ACCESSION AC005350
VERSION 93366562
KEYWORDS AC005350.1 GI:3366562
SOURCE HTS.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 78661)
AUTHORS Kimerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pitluk,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE Sequencing of human chromosome 5
JOURNAL unpublised
REFERENCE 2 (bases 1 to 78661)

AUTHORS	Ricke,D.O.
TITLE	Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System
JOURNAL REFERENCE	Unpublished 3 (bases 10 to 78661)
AUTHORS	Kimmel,M., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,R.M., Davis,C.A., Kadner,K., Miquel,T., Plickus,S., Pollard,M., Rojestki,H., Subramanian,S. and Martin,C.H.
TITLE	Direct Submission
JOURNAL	Submitted (1-JUL-1998) Human Genome Center, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, MS 74-157, Berkeley, CA 94720, U.S.A.
COMMENT	Sequence submitted by: DOE Joint Genome Institute. Location/Qualifiers
FEATURES	1..78661
source	/organism="Homo sapiens" /db_xref="taxon:9606" /map="sg" /clone="261j17" /chromosome="5" /note="LBNL H190" 476..656 /rpt_family="Alu" 2704..2839 /note="GRAIL 2 excellent exon, frame 0" 4504..4592 /rpt_family="MIR" 8089..8226 /rpt_family="Alu" complement(8516..9100) /rpt_family="MT2B2" 9398..10759 /rpt_family="L1" complement(10780..14299) /rpt_family="L1" complement(15044..15135) /rpt_family="MIR" complement(16072..16399) /rpt_family="LINE" 19038..19075 /note="(AT)19" /rpt_type=tandem /rpt_unit=AT complement(19076..19390) /rpt_family="Alu" 23689..24660 /rpt_family="L1" 24716..24995 /rpt_family="Alu" 25177..27648 /rpt_family="L1" 26615..26635 /note="(T)21" /rpt_type=tandem /rpt_unit=T complement(26618..26908) /rpt_family="Alu" complement(29199..29244) /note="GRAIL 2 excellent exon, frame 2" complement(30575..30784) /note="GRAIL 2 excellent exon, frame 0" 31846..31869 /note="(TA)12" /rpt_type=tandem /rpt_unit=TA complement(34387..34549) /rpt_family="MER3" 35029..35056 /note="(T)28" /rpt_type=tandem /rpt_unit=T complement(35039..35334)

Repeat	Repeat_Region	Repeat_Type	Repeat_Length	Repeat_Count	Repeat_Score	Repeat_DB	Repeat_Length	Repeat_Count	Repeat_Score	Repeat_DB
misc_feature	repeat_region	repeat_type=AC	36735..36764	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	39883..40023	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	41142..41292	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	41523..41837	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	45017..45258	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	46124..46204	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	47776..47817	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	51783..51857	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	55453..55706	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	57577..57597	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	66131..67648	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	69340..69369	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	70474..70725	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	72130..72130	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	73330..73522	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	75289..75289	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	75291..75291	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	77696..77815	1	76.5%	DB 11	78661	1	76.5%	DB 11
misc_feature	repeat_region	repeat_type=AC	14357 a 14357 c 14140 g 24597 t	1	76.5%	DB 11	78661	1	76.5%	DB 11

26 64143 TTAAAGTACCAATTAATGAGATCATCC 64176

```

RESULT 13
LOCUS AC006222
DEFINITION Homo sapiens, clone hRPK.12_A.1, complete sequence.
ACCESSION AC006222
MID 94159874
VERSION AC006222.1 GI:4159874
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 165643)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone hRPK.12_A.1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165643)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P.,
FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyne,S.,
Grant,G., Hagos,B., Hearford,A., Herena,L., Horton,L., Howland,J.C.,
Jones,C., Kann,L., Karatas,A., Lehoczy,J., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrum,J.,
Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J.,
Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,R., Peterson,K.,
Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramaniam,A.,
Tesiaye,S., Tortunella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE Direct Submission
JOURNAL * Submitted (13-DEC-1998) Whitehead Institute/MIT Center for Genome
REFERENCE Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS 3 (bases 1 to 165643)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeRellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyne,S., Gilbert,D., Grant,G.,
Hagos,B., Hearford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Liu,C., Locke,K., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McDonald,P.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramaniam,A.,
Tesiaye,S., Tortunella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-1999) Whitehead Institute/MIT Center for Genome
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 15, 1999 this sequence version replaced gi:4153849.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
source location/Qualifiers
1..165643
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK.12_A.1"
/clone_lib="RRC1-11 human BAC library"
repeat_region 111..2188
/rpt_family="L1PA12"
repeat_region 2185..2305
/rpt_family="L1M2"
repeat_region 2320..2427
/rpt_family="L1P"
repeat_region 2433..5941

```

```

repeat_region 111..2188
/rpt_family="L1PA13"
complement(6541..6645)
/rpt_family="MST0"
repeat_region 2185..2305
complement(6647..6753)
/rpt_family="MSTC"
repeat_region 2320..2427
complement(7652..7836)
/rpt_family="L2"
repeat_region 2433..5941
complement(8505..8669)
/rpt_family="MLT1J1"
repeat_region 8717..8743
/rpt_family="TTTA)n"
complement(8747..9035)
/rpt_family="L1M3A"
repeat_region 9048..9477
complement(9524..9585)
/rpt_family="L1PA7"
repeat_region 9556..9991
/rpt_family="MLT1J1"
repeat_region 10432..10716
/rpt_family="L1M3A"
repeat_region 10968..11016
/rpt_family="AT_rich"
repeat_region 11120..11152
/rpt_family="AT_rich"
repeat_region 11595..11970
/rpt_family="AT_rich"
repeat_region 12166..12264
/rpt_family="MLT1G"
repeat_region 12419..12687
/rpt_family="MLT1H"
repeat_region 12898..13042
/rpt_family="ALUS9"
complement(12898..13042)
/rpt_family="MIR"
repeat_region 13417..13508
/rpt_family="GA_rich"
repeat_region 13491..13849
/rpt_family="GGAAn"
repeat_region 13598..13653
/note="Single-stranded terminator coverage."
repeat_region 13839..13906
/rpt_family="GA_rich"
complement(14648..14707)
/rpt_family="L2"
repeat_region 16861..17001
/rpt_family="MLT1A1"
repeat_region 17385..17439
/rpt_family="MLT1I"
complement(17652..18126)
/rpt_family="MLT1D"
repeat_region 18170..18231
/rpt_family="MLT1H"
complement(21556..21881)
/rpt_family="L1MCS5"
repeat_region 22385..22592
/rpt_family="MIR"
repeat_region 22603..22668
/rpt_family="MER81"
repeat_region 22901..23033
/rpt_family="MER5B"
repeat_region 23204..23232
/rpt_family="TTTTC)n"
repeat_region 24016..24045
/rpt_family="TTTTC)n"
complement(24046..24373)
/rpt_family="L1PA9"
repeat_region 24621..24707
/rpt_family="AT_rich"
repeat_region 24764..24795
/rpt_family="TTTA)n"
complement(24796..25079)
/rpt_family="ALUSp"
repeat_region 26220..26243
/rpt_family="AT_rich"

```

```

repeat_region      26626..26675
                    /rpt_family="T1n"
repeat_region      complement(28534..28574)
                    /rpt_family="L2"
repeat_region      28669..28739
                    /rpt_family="AT_rich"
repeat_region      complement(29411..29497)
                    /rpt_family="MER53"
repeat_region      29498..29880
                    /rpt_family="MLT1A1"
repeat_region      complement(29881..29948)
                    /rpt_family="MER53"
repeat_region      complement(30188..30473)
                    /rpt_family="ALUSc"
repeat_region      complement(30674..31134)
                    /rpt_family="L2"
repeat_region      31136..31181
                    /rpt_family="AT_rich"
repeat_region      complement(31200..31404)
                    /rpt_family="MLT1K"
repeat_region      32833..32970
                    /rpt_family="MIR"
repeat_region      complement(33783..33864)
                    /rpt_family="L1ME3"
repeat_region      complement(34307..34421)
                    /rpt_family="MER5A"
repeat_region      complement(34525..34813)
                    /rpt_family="L1ME"
repeat_region      complement(35163..35330)
                    /rpt_family="MER45"
repeat_region      complement(35581..35865)
                    /rpt_family="ALUSx"
repeat_region      36066..36196
                    /rpt_family="MIR"
repeat_region      36393..36554
                    /rpt_family="MIR"
repeat_region      36655..37111
                    /rpt_family="L2"
repeat_region      37139..37286
                    /rpt_family="MIR"
repeat_region      37287..37595
                    /rpt_family="ALUSg"
repeat_region      37596..37640
                    /rpt_family="MTR"
repeat_region      complement(37777..38164)
                    /rpt_family="L1R47A"
repeat_region      38165..38259
                    /rpt_family="L2"
repeat_region      38295..38329
                    /rpt_family="(TTTA)n"
repeat_region      complement(38330..38611)
                    /rpt_family="ALUSx"
repeat_region      38688..38954
                    /rpt_family="MLT1J2"
repeat_region      complement(40069..40207)
                    /rpt_family="L2"
repeat_region      complement(40290..40607)
                    /rpt_family="L2"
repeat_region      40883..41336

Query Match      60.6%; Score 21.2; DB 11; Length 165643;
Best Local Similarity 88.5%; Pred. No. 20;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY      7 CTTAGGCAATCTAAAGTCACTATCAT 32
          |||||
Db      7361 CTTAGGCAATCTAAAGTCACTTCC 7386

```

```

RESULT 14
LOCUS      KTU09810      15532 bp      DNA      circular      INV      22-JUL-1998
DEFINITION      Katharina tunicata mitochondrial genome, complete sequence.

```

```

ACCESSION      U09810
MID      9537273
VERSION      U09810.1 GI:557273
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  1 (bases 1 to 15532)
  black chiton.
  Eukaryota; Metazoa; Mollusca; Polyplacophora; Neoloricata;
  Ischnochitonida; Mopallidae; Katharina.
  2 (bases 1 to 15532)
  Boore, J.L. and Brown, W.M.
  Complete DNA sequence of the mitochondrial genome of the black
  chiton, Katharina tunicata
  Genetics 138 (2), 423-443 (1994)
JOURNAL
MEDLINE
  95129806
  2 (bases 1 to 15532)
  Boore, J.L.
  Direct Submission
  Submitted (17-MAY-1994) Jeffrey L. Boore, Cell Biology and
  Neuroanatomy, University of Minnesota, 321 Church St. SE,
  Minneapolis, MN 55455, USA
FEATURES
  source
  1..1532
  location/Qualifiers
  /organism="Katharina tunicata"
  /mitochondrion
  /db_xref="taxon:34587"
  1..1548
  /note="followed by 'r' as an abbreviated stop codon"
  /codon_start=1
  /transl_table=5
  /codon=(seq:"tga", aa:Trp)
  /product="cytochrome oxidase subunit I"
  /protein_id="AAC48362.1"
  /db_xref="pid:9537274"
  /transl_table=5
  /translation="MRWIFSTNHRKDIGTLLTIFGIAGLVGTALSTLIRAEIGPGAL
  LGDDOLYVIVAHAFVAFVPMVPMGGMVPLMLGVDPDAFRLNMSFWL
  PPAICLLASGAVESGAGTGWVPPVPLAGNVGAGSVDAIFSLHAGVSSIIAGV
  EFTTVNRSRSGMOLERPLFVMSKKTATLILSLSTPLAGITMLDRPNNSFD
  PAGGDPYIOHLEFHFGRHPEYVITLLFGFGLSHIVHSSKETFTGLNITAMLA
  IGLGFIYAHMHPVGVGDVDTRAFVATMIAVPGIKIFSWLATVIGARIGETP
  MMWALGFLEFTEVGLTGLIVLSNSLIDIMLDSYVVAHFVYVSMGVAFALFGAFY
  WYPLNTGLSLHERMTSHVYVMEFGLVNLTFPPHGLSGMPPRYSDDPCYIKWVY
  SSIGLSIFVAVLFPNFIWVESIVSQREVIVSSHLSALEMDNSIPLGFTDLMKGL
  FVNLITL"
  1619..1682
  /note="codon recognized: GAC"
  /product="tRNA-Asp"
  /anticodon="(pos:1650..1652,aa:Asp)
  2107..2796
  /codon_start=1
  /transl_table=5
  /codon=(seq:"tga", aa:Trp)
  /product="cytochrome oxidase subunit II"
  /protein_id="AAC48365.1"
  /db_xref="pid:9537277"
  /db_xref="gi:557277"
  /translation="MAFWSQGFQDQASPIEQLIFRDHMLILIMISLSGAVS
  LMNSFSLRSTLESQETETVITLPAVVLTLAPPSQLTLVLEDELPALTIVYGH
  QWYWSYEDSLNLEFDSYMSLDELDGDRRLLEVDRSVYPMKTKRVLTADVL
  HSWTVPSLGVKADAVPGRNLQSFANYPGVFGQCSBICGANSFMRITVEVDSSS
  FIKWIMFNGEA"
  2872..3033
  /codon_start=1
  /transl_except="(pos:3001..3003,aa:Trp)
  /transl_table=5
  /product="ATPase subunit 8"
  /protein_id="AAC48364.1"
  /db_xref="pid:g3335323"
  /db_xref="gi:3335323"
  /translation="MPQLAPMNWIFLLLFWSNVFCIGVCVMMWSSSSEFLSKSVGN
  PFYFKWVW"
  3037..3729
  /codon_start=1

```

/transl_table=5
/product="ATPase subunit 6"
/protein_id="AAC48363.1"
/db_xref="PID:9557275"
/db_xref="GI:557275"
/translation="MADDFSSFDNNFTLFSFYLIMVLGLAVLISQNMWFSKSR
FYSLPFPKNGMGQVWATKSNISGFINLWVSLILVNLGSLIPYVSLSSHS
FAPFCGLPFLSLFSGDLVSEVSHLPSGAGILNPPLVIVETYSISVRPTLSY
RLANMSAGHILIGLSVGLFCIPILVGLLLEVOVFYFLFELGSLIQATIRS
LVLVYSDDHPY"
/complement(3758..3822)
/note="codon recognized: TTC"
/product="tRNA-Phe"
/anticodon="(pos:3789..3791,aa:Phe)
/complement(3821..3536)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 5"
/protein_id="AAC48366.1"
/db_xref="PID:9557278"
/db_xref="GI:557278"
/translation="MKMFLSTQATVILLSHSLSESTAMVTLTKIILLECFFN
ENSVITLPIINVTALFSAIVCIYASVMIFFKNTMSODIFIKRFTWMLVLSM
NLIFPSIALTLGMDGIGVSRCLVYONKYSGLAGMTAFMRNIGDAIILATS
VTSQGHFNLPMSTPTMTITSMIILAGTRSAQCPFCSMPLAAMAAPPVSAIVH
SSVLVAGVEVLIDVYFPLNSHLSTSLMITSWVMAGISANFETDKRIILALT
LSQGVNMLSVSLGFLFSLALFHLTHAMFKALFLPCAGNIHSHNNQDILKSHMT
QMPFSTCFNLANLGCFFPMAGYKDVIIEMMFNQNNFLISLNNFLATILASY
SARSLFPMKNTQSSVSTSDENSIAPLTLAGAITSGAVFSMLILPTSLPL
ISMNKRMATEPATLGAATITKLEIKLISHTSIINFFEMEMNINLSNMTRKSSL
YTGITPKSLDGMETIGGBGLNKRNTMTTITMOKNOSNLNPLFISLILSMPLTSV
TMTT"
/complement(5537..5601)
/note="codon recognized: CAC"
/product="tRNA-His"
/anticodon="(pos:5568..5570,aa:His)
/complement(5597..6928)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_except="(pos:complement(6788..6790),aa:Trp)
/transl_table=5
/product="NADH dehydrogenase subunit 4"
/protein_id="AAC48367.1"
/db_xref="PID:9335324"
/db_xref="GI:3335324"
/translation="MSAFSTVMTKSKMNYKMGMSIHFMEFNTLNEFSMPNPKIN
YEMSDNESVIAMLSCMITGLMINSYKILKNMLVSEFLNMTINFTIIMVETOK
SISPLIEFEASLIPRLIILIMGQYPERQAGVMMIVITIGALPRLINFTYSHN
AHLNLLMSLPIPYQAMISFHWLITILVPLVLPITYSFHMLPKAHEAPVAGSMI
LAALLKLGCGINTINNNTIKYFNVLNFIILGLVGGVSLICLRQSMKALAY
SSVGHGMLAGLISSEFGLKMLMLMIAHGLSSGIFSNNMYEKSGRSFLFK
GFLSVPTFSKSLFIMCSINNAPIINLSEALMISVSIKFLMLLMAFVSA
VYTLFTNTHGGLISSFVNPFNNFKPLVYVIFLHMIPADLLILLSTTC"
/complement(6940..7242)
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 4L"
/protein_id="AAC48368.1"
/db_xref="PID:9557280"
/db_xref="GI:557280"
/translation="MANFNPFLTGLALSSISALMTICLQKHLNLTLLALEITLWVSI
FLLVLSLSNLSNGELMIFVLLTLAASEASIGLMLVILLRAHNDVYSISIRKC"
/complement(7312..7312)
/note="codon recognized: ACA"
/product="tRNA-Thr"
/anticodon="(pos:7275..7277,aa:Thr)
/complement(7312..7378)
/note="codon recognized: TCA"
/product="tRNA-Ser"
/anticodon="(pos:7348..7350,aa:Ser)
/complement(7381..8520)
/codon_start=1

/transl_except="(pos:complement(8176..8178),aa:Trp)
/transl_table=5
/product="cytochrome b"
/protein_id="AAC48371.1"
/db_xref="PID:9335325"
/db_xref="GI:3335325"
/translation="MKPIRSQHVPEKILNGSLIDPSPENSLTMMWFGSLGLCLIT
OITGLFLAHYATNITELARSSVAHITRDVNYGCLTAAHNGSLFETICYPHIGG
MYGSEFNHRTWVGVILFLVTMGTAIGVLPWQMSFWGATVITLMSXAPIGTGT
MYQWINGCFVADNATLTFPESFHYLAFLIAASVHLILFELHGSNNPLGINDSEK
VTIHSEFTLKDLYGVVVFLLVLELDELNLAPNDIPLANPLVPTPHQPEMYEL
FAVALISPKRIGGVGLMSIVLEFVPLINRGKRTISFYVPTQILVWSLVSVEL
LTMWGAQPVDPEDPILGOLITTFYFLFLMIPRTKMMDMTINF"
/complement(8513..9019)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 6"
/protein_id="AAC48369.1"
/db_xref="PID:9557281"
/db_xref="GI:557281"
/translation="MFLTMSLIFSCAFSLPLTQPIISGLLPLNSPLISILIFG
SSWFSFLFLVIGLGLVMEAYVTALPMLMEKRLIQVFPVSTVFMLMLTSDP
IGTLEROTKMFIMNPFHMLGSLSPFNLIIVSLALILEFVLICVYKICVFSNG
PLRPYKXA"
/complement(9024..9089)
/note="codon recognized: CCA"
/product="tRNA-Pro"
/anticodon="(pos:9054..9056,aa:Pro)
/complement(9082..10032)
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 1"
/protein_id="AAC48370.1"
/db_xref="PID:9557282"
/db_xref="GI:557282"
/translation="MFLSFSMMVAYISILLAVFTLLERKGIYQFRPKVYSI
FGLAQPLADAKLFLKEENNYNSNKLPYTAAPLSLISLCMSLWSDPTWTFNF
GVLFELTQSLALSVYGTMIACWASNSKALLGALAAVNAQTSYEVELVLISALSS
SLNINQISSHQULVWNSILMTPLIMMIVVCLALMTNAPDPAFAGESELVSGENIERG
SGSFALLFLAEVYASILFNSMNTTILFAGAGSWILTNLEFVYKVTFAFYIVWGTFF
PRLRYDILMKLTWISFLPVSIMTILMKMLSKFNTVITLLS"
/complement(10033..10101)
/note="codon recognized: TTA"
/product="tRNA-Leu"
/anticodon="(pos:10068..10070,aa:Leu)
/complement(10101..10165)
/note="codon recognized: CTA"
/product="tRNA-Leu"
/anticodon="(pos:10132..10134,aa:Leu)
/complement(10166..11440)
/note="The exact endpoints are uncertain"

Query Match 60.6%; Score 21.2; DB 37; Length 15532;
Best Local Similarity 76.5%; Pred. No. 34;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 10306 TTTAAGCTTAGGCATTCTAAGTCACTATCAATCC 34
DB 10306 TTTAAGCTTCGCACTTAAGTACTTAACTTAAATCC 10339

RESULT 15
SLH4K 1673 bp DNA INV 12-SEP-1993
LOCUS STYLONGYCHIA lemnae histone H4 gene (clone H4K).
DEFINITION X16018
ACCESSION X16018
NID 910141
VERSION X16018.1 GI:10141
KEYWORDS histone; histone H4.
SOURCE Stylongychia lemnae.
ORGANISM Stylongychia lemnae
Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichidae;

```

REFERENCE      Oxytrichidae; Stylonychia.
AUTHORS        1 (bases 1 to 1673)
TITLE          Direct Submission
JOURNAL        Submitted (03-AUG-1989) Wefes I.,
               Medizinische-Naturwissenschaftliches Forschungszentrum,
               Universitaet Tuebingen, Ob dem Himmelreich 7, 7400 Tuebingen, F R G
REFERENCE      2 (bases 1 to 1673)
AUTHORS        Wefes, I. and Lipps, H.J.
TITLE          The two macronuclear histone H4 genes of the hypotrichous ciliate
               Stylonychia lemnae
JOURNAL        DNA Seq. 1 (1), 25-32 (1990)
MEDLINE        92119316
COMMENT        Data kindly reviewed (19-NOV-1990) by Lipps H.J.
FEATURES       Location/Qualifiers
               source
               1..1673
               /organism="Stylonychia lemnae"
               /strain="EK16Sp"
               /db_xref="taxon:5949"
               /clone="H4K"
               1..20
               /note="telomeric sequence"
               misc_feature
               365..401
               /note="ARS sequence"
               misc_feature
               785..821
               /note="ARS sequence"
               misc_feature
               complement(1014..1024)
               /note="ARS sequence"
               misc_feature
               1072..1103
               /note="ARS sequence"
               promoter
               1117..1121
               /note="CAAT-box"
               promoter
               1189..1195
               /note="TATA-box"
               CDS
               1208..1522
               /note="histone H4"
               /codon_start=1
               /transl_table=6
               /protein_id="CAA34151.1"
               /db_xref="PID:910142"
               /db_xref="GI:10142"
               /translation="MAGRGKVGKGGKVGAKRHTKSLKETIMGITKPAIRRLARRG
               VKRISLIYEFRNVLRSFLENVIRDSVTYTHAKRKYTALDVTYALKRQRTLIGF
               GG"
               misc_feature
               1557..1588
               /note="ARS sequence"
               misc_feature
               1576..1598
               /note="ARS sequence"
               misc_feature
               1654..1673
               /note="telomeric sequence"
BASE COUNT    606 a 276 c 266 g 525 t
ORIGIN
Query Match      59.4%; Score 20.8; DB 36; Length 1673;
Best Local Similarity 78.1%; Pred. No.79;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY      1 TTTAAGCTTAGGCATCTTAAGTCACTATCAT 32
        ||||| ||||| ||||| |||||
DB      1163 TTTAAGCGACTGCATTTTAAGTCTCTATAT 1194

```

Search completed: September 13, 1999, 15:55:23
Job time: 4551 sec

CC Isolation of the phosphatidyl inositol (PtdIns) 3-kinase cDNA. The
CC amplified sequence was placed under the regulatory control of the nmt
CC promoter. These sequences were transformed into Schizosaccharomyces
CC pombe cells in an embodiment of the invention. In the presence of
CC thiamine the promoter is inactive and the cells carrying the PtdIns
CC plasmids grow as the parental strain. In the absence of thiamine the
CC nmt promoter functions and the PtdIns is induced. PtdIns activity is
CC substantially increased under these conditions. Cells containing
CC constructs such as this, are useful in assays for detecting compounds
CC involved in cell growth regulation. It is also used as the basis for
CC detecting compounds for treating cancers and the formation of blood
CC vessel plaques.

Sequence 35 BP; 10 A; 9 C; 4 G; 12 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 2e-06;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

07 1 TTTAGCTTAAAGCTTAAAGCTATCATCC 35
DB 1 TTTAGCTTAAAGCTTAAAGCTATCATCC 35

RESULT 3
051155/c
ID 051155 standard; cDNA; 3207 BP.

AC 051155;
DE 12-APR-1994 (first entry)

DT p110 cDNA.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KM antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KM 85.

FT cds Location/Qualifiers

FT 1..3207

FT /tag= a

FT /note= "PI3- kinase p110"

PD WO9321328-A.

PN 28-OCT-1993.

PR 13-APR-1993; G00761.

PR 13-APR-1993; GB-008135.

PA (LUDWIG-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI Parker PJ, Volinia S, Waterfield MD;

DR P-PSDB; R43341.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PS Claim 7; fig 9; 146pp. English.

CC Southern blot analysis was performed using a bovine cDNA probe conty.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC prophylaxis or therapy. Platelet or neutrophil activity or blood

CC growth levels can be controlled using the kinase.

CC See also 051156, 059012-23 and 057522-3.

CC Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

07 10 AGGATTTAAAGTCACTATCATCC 35

|||||

DB 110 AGGATTTAAAGTCACTATCATCC 85

RESULT 4

ID 057012/c

AC 057012 standard; cDNA to mRNA; 3498 BP.

DE 31-AUG-1994 (first entry)

DT PtdIns 3-kinase 110 KD catalytic subunit cDNA.

KW 110 KD catalytic subunit; phosphatidyl inositol 3-kinase;

KM transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;

KW PtdIns 3-kinase; assay; detection; cell growth; regulation; cancer;

OS blood vessel plaques; 85.

FT cds Location/Qualifiers

FT 1..3207

FT /tag= a

FT /product= p110

PN WO9403609-A.

PD 17-FEB-1994.

PR 05-AUG-1993; G01651.

PR 05-AUG-1993; GB-016654.

PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.

PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;

DR WPI; 94-065697/08.

DR P-PSDB; R46294.

PT Eukaryotic cells transformed with mammalian phospholipid or

PT protein kinase DNA - useful in assays for compounds involved in

PT cell growth regulation and for treating cancers

PS Disclosure: fig 1; 71pp. English.

CC This sequence encodes the 110 KD catalytic subunit of the phosphatidyl

CC inositol (PtdIns) 3-kinase. This sequence was transformed into

CC Schizosaccharomyces pombe cells under the regulatory control of the

CC nmt promoter in an embodiment of the invention. In the presence of

CC thiamine the promoter is inactive and the cells carrying the PtdIns

CC catalytic subunit plasmid grow as the parental strain. In the absence

CC of thiamine the nmt promoter functions and the PtdIns 3-kinase

CC catalytic subunit is induced. PtdIns activity is substantially

CC increased under these conditions. Cells containing constructs such as

CC this, are useful in assays for detecting compounds involved in cell

CC growth regulation. It is also used as the basis for detecting

CC compounds for treating cancers and the formation of blood vessel

CC plaques.

Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 74.3%; Score 26; DB 1; Length 3498;

Best Local Similarity 100.0%; Pred. No. 0.024;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

07 10 AGGATTTAAAGTCACTATCATCC 35

|||||

DB 110 AGGATTTAAAGTCACTATCATCC 85

RESULT 5

ID 051156/c

AC 051156 standard; cDNA; 3412 BP.

DE 12-APR-1994 (first entry)

DT Human p110 cDNA.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KM antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KW ds.

OS Human.

FT key Location/Qualifiers

FT cds 1..3207

FT /tag= a

FT /note= "PI3- kinase p110"

PN WO9321328-A.

PD 28-OCT-1993.

PR 13-APR-1993; G00761.

```

PR 13-APR-1992: GB-008135
PA (LUDW.) LODWIG INST CANCER RES.
PI Dhand R, Fry MO, Gout I, Hiles ID, Otsu M, Panayotou G;
PI Parker PJ, Volinia S, Waterfield MD;
DR WPI: 93-351738/44.
P-PSDB: R43342.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Claim 7; Fig 16; 146pp; English.
CC Southern blot analysis was performed using a bovine cDNA probe contg.
CC a fragment of a p13-kinase-encoding sequence and human cDNA isolated
CC from a cDNA library constructed from mRNA isolated from the human
CC cell line K562. Positive clones were sequenced to give the human
CC p13 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with p13 kinase activity and is useful for
CC screening for (anti)agonists of p13 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also Q51155 and Q57522-3.
SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 69.7%; Score 24.4; DB 1; Length 3412;
Best Local Similarity 96.2%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 10 AGGATTCTTAAGTCATCATTC 35
DB 110 AGGATTCTTAAGTCATCATTC 85

RESULT 6
O77252/C
ID O77252 standard; DNA: 314 BP.
AC O77252;
DT 23-SEP-1994 (first entry)
DE Human genome fragment (Preferred).
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
OS Homo sapiens.
PN W03401548-A.
PD 20-JAN-1994.
PF 13-JUL-1993: G01467.
PR 13-JUL-1992: GB-014857.
PA (MEDT.) MEDICAL RES COUNCIL.
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D.
PI Sibson DR, Starkey M;
DR WPI: 94-035056/04.
PT New nucleic acid fragment encoding gene products - can be used
PT for genetic analysis and mapping
PT Claim 1; Page 429; 616pp; English.
CC Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC placenta or bone marrow comprise any of: (A) a sequence
CC selected from (G76401-G7613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
CC Preferred sequences exhibit no more than 90% homology to a human
CC sequence known per se.
SV Sequence 314 BP; 96 A; 65 C; 58 G; 95 T;

Query Match 56.6%; Score 19.8; DB 1; Length 314;
Best Local Similarity 77.4%; Pred. No. 7.3;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 TTAAGCTTAGGATTCTTAAGTCATCATTC 32
DB 138 TAAAGCTTAGGATTCTTCAAAAGTCATTCAT 108

```

```

RESULT 7
V52190
ID V52190 standard; DNA: 10669 BP.
AC V52190;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:57.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN W09816931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI: 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1; Page 504-510; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 10669 BP; 3245 A; 1967 C; 2353 G; 3104 T;

Query Match 56.6%; Score 19.8; DB 1; Length 10669;
Best Local Similarity 77.4%; Pred. No. 12;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 4 AAGCTTAGGATTCTTAAGTCATTCATTC 34
DB 2033 AAAGCTTAGGATTCTTAAGTCATTCATTC 2063

RESULT 8
T62680
ID T62680 standard; DNA: 1196 BP.
AC T62680;
DT 28-MAY-1997 (first entry)
DE Baboon SP-10 CDNA encoding acrosomal sperm antigen 10.
KW SP-10; spermatogenesis; acrosomal sperm antigen; membrane;
KW contraceptive vaccine; testis-specific; differentiation; ss.
OS Papio papio.
PT Key
PT misc_signal
PT Location/Qualifiers
FT tag= a
FT /note= "5' consensus sequence flanking ATG start"
FT cds 72..929
FT /tag= b
FT signal_peptide 72..122

```

```

FT      mat_peptide      /tag- c
FT      123..926
FT      /tag- d
FT      misc_feature      583..684
FT      /tag- e
FT      /note- "alternately spliced out"
FT      misc_signal      1000..1004
FT      /tag- f
FT      /note- "mRNA degradation consensus sequence"
FT      poly_a_signal      1166..1171
FT      /tag- g
PN      US5602005-A.
PD      11-FEB-1997.
PF      03-MAR-1989; 318551.
PR      03-MAR-1989; US-318551.
PR      16-FEB-1990; US-481491.
PR      27-MAR-1992; US-858798.
PR      18-AUG-1994; US-292045.
PA      (UYV-) UNIV VIRGINIA PATENTS FOUND.
PI      Herr JC, Wright RM;
DR      MPI: 97-131798/12.
DR      P-PSDB; W14465.
PT      Intra-acrosomal primate sperm antigen - for contraceptive vaccine
PS      Prodn.
CC      Example 14; Fig 16; 70pp: English.
CC      The present sequence is SP-10 cDNA isolated from a baboon testis cDNA
CC      library. Alternative splicing is exhibited within the open reading frame,
CC      with a 102 bp internal deletion. The flanking sequence of this deletion
CC      encodes the 5' GTG-GAG 3' consensus splice sequence characteristic of an
CC      intron. The SP-10 protein is an intra-acrosomal primate sperm antigen
CC      that remains associated with the primate sperm after the acrosome
CC      reaction and is bound by monoclonal antibody produced by cell line ATCC
CC      HB10039. The antigen can be used for the production of contraceptive
CC      vaccines.
SQ      Sequence      1196 BP;      307 A;      287 C;      259 G;      343 T;

Query Match      56.0%; Score 19.6; DB 1; Length 1196;
Best Local Similarity 73.5%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      1 TTTAGCTTAGCATCTCTAAGTCACTATCATCC 34
      ||| || ||||| || ||||| ||||| |||||
DB      949 TTTGACTCAGCAGTAATAAGTCTCTATCATTC 982

RESULT 9
ID      T62681 standard: DNA; 1196 BP.
AC      T62681;
DT      28-MAY-1997 (first entry)
DE      Macaque SP-10 cDNA encoding acrosomal sperm antigen 10.
KW      SP-10; spermatogenesis; acrosomal sperm antigen; membrane;
KW      contraceptive vaccine; testis-specific; differentiation; ss.
OS      Macaca fascicularis.
PI      Key
PI      Location/Qualifiers
FT      misc_signal      4
FT      /tag- a
FT      /note- "major transcriptional start site"
FT      65..71
FT      /tag- b
FT      /note- "5' consensus sequence flanking ATG start"
FT      72..929
FT      /tag- c
FT      /tag- d
FT      /tag- e
FT      /tag- f
FT      /tag- g
FT      /note- "alternately spliced out"
FT      1000..1004
FT      /tag- g

```

```

FT      poly_a_signal      /note- "mRNA degradation consensus sequence"
FT      1166..1171
FT      /tag- h
PN      US5602005-A.
PD      11-FEB-1997.
PF      03-MAR-1989; 318551.
PR      03-MAR-1989; US-318551.
PR      16-FEB-1990; US-481491.
PR      27-MAR-1992; US-858798.
PR      18-AUG-1994; US-292045.
PA      (UYV-) UNIV VIRGINIA PATENTS FOUND.
PI      Herr JC, Wright RM;
DR      MPI: 97-131798/12.
DR      P-PSDB; W14466.
PT      Intra-acrosomal primate sperm antigen - for contraceptive vaccine
PS      Prodn.
CC      Example 14; Fig 16; 70pp: English.
CC      The present sequence is SP-10 cDNA isolated from a macaque testis cDNA
CC      library. Alternative splicing is exhibited within the open reading frame,
CC      with a 102 bp internal deletion. The flanking sequence of this deletion
CC      encodes the 5' GTG-GAG 3' consensus splice sequence characteristic of an
CC      intron. The SP-10 protein is an intra-acrosomal primate sperm antigen
CC      that remains associated with the primate sperm after the acrosome
CC      reaction and is bound by monoclonal antibody produced by cell line ATCC
CC      HB10039. The antigen can be used for the production of contraceptive
CC      vaccines.
SQ      Sequence      1196 BP;      304 A;      283 C;      261 G;      348 T;

Query Match      56.0%; Score 19.6; DB 1; Length 1196;
Best Local Similarity 73.5%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      1 TTTAGCTTAGCATCTCTAAGTCACTATCATCC 34
      ||| || ||||| || ||||| ||||| |||||
DB      949 TTTGACTCAGCAGTAATAAGTCTCTATCATTC 982

RESULT 10
ID      V62749/c
ID      V62749;
AC      15-FEB-1999 (first entry)
DE      Human secreted protein clone er418_5 cDNA.
KW      Secreted protein; human; er418_5; ds.
OS      Homo sapiens.
PI      Key
PI      Location/Qualifiers
FT      CDS      571..3306
FT      /tag- a
PN      M09846757-A2.
PD      22-OCT-1998.
PF      14-APR-1998; U07999.
PR      13-APR-1998; US-059487.
PR      15-APR-1997; US-843374.
PA      (GENV) GENETICS INST INC.
PI      Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI      Racine JA, Spaulding V, Treacy M;
DR      MPI: 98-568731/48.
DR      P-PSDB; W74723.
PT      New polynucleotide(s) encoding secreted human proteins - are derived
PT      from, e.g. human foetal brain or foetal kidney cDNA libraries,
PT      potentially useful as, e.g. vaccines or thrombolytic agents
PS      Claim 19; page 76-78; 120pp: English.
CC      Full-length cDNA clone er418_5 includes an open reading frame
CC      encoding a human secreted protein (see W74723). It was isolated
CC      from a human foetal brain cDNA library using methods which are
CC      selective for cDNAs encoding secreted proteins, or was identified
CC      as encoding a secreted or transmembrane protein on the basis of
CC      computer analysis of the amino acid sequence of the encoded protein.
CC      It shows some similarity to database sequences. The invention
CC      provides polynucleotides (see V62746-55) from human foetal brain,
CC      adult testis, adult brain, adult kidney and foetal kidney (all
CC      deposited as composite clone ATCC 98404), which encode human

```


CC secreted proteins (see W74720-29). The polynucleotides and
 CC proteins are predicted to have biological activities which would
 CC make them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data are
 CC given. Suggested activities include nutritional activity, immune
 CC stimulating (e.g. as vaccines) or suppressing activity.
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity,
 CC haemostatic and thrombolytic activity, receptor/ligand activity,
 CC antiinflammatory activity, cadherin/tumour invasion suppressor
 CC activity, and tumour inhibition activity. The polynucleotides are
 CC also stated to be useful for gene therapy, and can be used in
 CC recombinant production of the polypeptides.
 SQ Sequence 4130 BP; 1113 A; 1051 C; 1024 G; 936 T;

Query Match 56.0%; Score 19.6; DB 1; Length 4130;
 Best Local Similarity 73.5%; Pred. No. 12;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 2 TTAAGCTTAGCATTCTAAAGTCATCATGCC 35
 DB 794 TTAATCTTAGCATTTTCAAGGACCTCCTTCC 761

RESULT 11
 V78144
 ID V78144 standard; DNA; 400 BP.
 AC V78144;
 DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #3833.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 OS Staphylococcus aureus.
 PN EP-786519-A2.
 PD 30-JUL-1997.
 PR 07-JAN-1997; 100117.
 PR 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DT WPI: 97-374922/35.
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 PS Claim 1; Page 2708; 3271bp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 400 BP; 114 A; 77 C; 67 G; 139 T;

Query Match 54.9%; Score 19.2; DB 1; Length 400;
 Best Local Similarity 75.0%; Pred. No. 14;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 3 TAAAGCTTAGCATTCTAAAGTCATCATGCC 34
 DB 251 TATGCTTAGCAAGCATTAATATCATCTGTACGCC 282

RESULT 12
 V78616
 ID V78616 standard; DNA; 344 BP.
 AC V78616;
 DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #4305.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 OS Staphylococcus aureus.
 PN EP-786519-A2.
 PD 30-JUL-1997.
 PR 07-JAN-1997; 100117.
 PR 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DT WPI: 97-374922/35.
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 PS Claim 1; Page 2911; 3271bp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 344 BP; 92 A; 60 C; 73 G; 116 T;

Query Match 54.9%; Score 19.2; DB 1; Length 344;
 Best Local Similarity 75.0%; Pred. No. 13;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 3 TAAAGCTTAGCATTCTAAAGTCATCATGCC 34
 DB 31 TATGCTTAGCAAGCATTAATATCATCTGTACGCC 62

RESULT 13
 V74342/C
 ID V74342 standard; DNA; 13856 BP.
 AC V74342;
 DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #31.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 OS Staphylococcus aureus.
 PN Key
 DT misc_feature 1, 60
 FT Location/Qualifiers
 FT /tag= a
 FT /note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence.
 FT misc-feature
 FT 1801. 1860
 FT /tag- b
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 3601. 3660
 FT /tag- c
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 5401. 5460
 FT /tag- d
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 7201. 7260
 FT /tag- e
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 9001. 9060
 FT /tag- f
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 10801. 10860
 FT /tag- g
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 12601. 12660
 FT /tag- h
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 308-315. 3271pp: English.
 PS Claim 1, Page 308-315. 3271pp: English.
 CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the *S. aureus* DNA sequences allows putative functions to be assigned so
 CC that protein encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against *S. aureus* infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating

CC homologues of any of the *S. aureus* DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 13856 BP; 4218 A; 2598 C; 2085 G; 4467 T;
 Query Match 54.3%; Score 19.2; DB 1; Length 13856;
 Best Local Similarity 75.0%; Pred. No. 21;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 3 TAACCTTAGGCATCTTAAGCTATCATCC 34
 DB 8125 TATCCTTAGGCATCTTAAGCTATCATCC 8094
 RESULT 14
 X13998/C
 ID X13998 standard; DNA; 2821 BP.
 AC X13998;
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHPD 984 gene.
 KW GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
 OS Helicobacter pylori.
 FH Key location/Qualifiers
 FT CDS 58..2772
 FT /tag- a
 FT W09843478-A1.
 PD 08-OCT-1998.
 PF 01-APR-1998; U06371.
 PR 29-JUL-1997; US-902615.
 PR 01-APR-1997; US-833457.
 PR 24-JUN-1997; US-881227.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PI A1-Garawi A, Kleantchous H, Miller C, Oomen RP, Tomb J;
 DR WPI: 98-542293/46.
 DR P-PSDB: W98279.
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 PS Claim 1; Page 317-321. 2054pp: English.
 CC This sequence represents a polynucleotide of the invention. It was
 CC isolated from Helicobacter pylori and encodes a H.pylori GHPD protein.
 CC The polypeptides can be used for preventing or treating Helicobacter
 CC infections, and gastroduodenal diseases associated with these
 CC infections, including acute, chronic, and atrophic gastritis, and peptic
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
 CC for the production of antibodies. The products can also be used for
 CC detection and diagnosis
 SQ Sequence 2821 BP; 918 A; 480 C; 663 G; 760 T;
 Query Match 54.3%; Score 19; DB 1; Length 2821;
 Best Local Similarity 81.5%; Pred. No. 21;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 6 GCTTAGGCATCTTAAGCTATCAT 32
 DB 2118 GCTTAGGCATCTTAAGCTATCAT 2092
 RESULT 15
 T24992
 ID T24992 standard; cDNA to mRNA; 240 BP.
 AC T24992;
 DT 07-NOV-1996 (first entry)
 DE Human gene signature HUGS07114.
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 OS cell typing; abnormal cell function; ss.
 PN W09514772-A1.
 PD 01-JUN-1995.

PF 11-NOV-1994: J01916.
 PR 12-NOV-1993: JP-355504.
 PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 FI Matsubara K, Okubo K;
 DR WPI: 95-206931/27.
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 PS Claim 1: Page 1744; 2245pp; Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in T19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 SQ Sequence 240 BP; 73 A; 41 C; 49 G; 72 T;

Query Match 53.7%: Score 18.8; DB 1; Length 240;
 Best Local Similarity 76.7%: Pred. No. 19;
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 OY 2 TTAAGCTTAGGCATCTTAAGTCACTATCA 31
 ||||| ||||| ||||| |||||
 Db 132 TTATGCTAGACATTTTAAGTAATTATGA 161

Search completed: September 13, 1999, 15:59:31
 Job time: 4798 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:55 ; Search time 1694.61 Seconds

(without alignments)
40.740 Million cell updates/sec

Title: US-09-325-095-17

Perfect score: 35
Sequence: 1 TTTAAGCTTAGGCACTTCTAAAGTCATCATCACC 35

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24.4	69.7	417	29	AA190409
2	20.8	59.4	307	36	AA610665
3	20.4	58.3	593	39	AA887564
4	20.4	58.3	389	44	A1265938
5	20.2	57.7	541	48	A1543543
6	19.8	56.6	314	20	221049
7	19.8	56.6	595	43	A1161953
8	19.8	56.6	557	43	A1162696
9	19.8	56.6	491	43	A1165066
10	19.8	56.6	842	43	A1188280
11	19.6	56.0	586	35	299424
12	19.6	56.0	605	41	A1050466
13	19.6	56.0	749	44	A1314059
14	19.6	56.0	620	53	HSM002526
15	19.6	56.0	762	53	HSM002976
16	19.4	55.4	372	20	T26601
17	19.2	54.9	442	25	N91396
18	19.2	54.9	460	27	AA015969
19	19.2	54.9	503	30	AA197781
20	19.2	54.9	390	34	AA528623
21	19.2	54.9	133	41	AU006843
22	19.2	54.9	633	47	A1518990
23	19.2	54.9	422	49	A1628722
24	19.2	54.9	515	22	R56494
25	19.2	54.3	427	24	N29688
26	19.2	54.3	508	24	N29717
27	19.2	54.3	597	26	W73757
28	19.2	54.3	493	30	AA242870
29	19.2	54.3	365	33	AA414436
30	19.2	54.3	418	33	AA416936
31	19.2	54.3	464	40	AA905139
32	19.2	54.3	363	19	AA905280
33	19.2	54.3	359	41	A1061048
34	19.2	54.3	349	43	A1220668
35	19.2	54.3	268	47	A1467067
36	19.2	54.3	376	47	A1509866
37	18.8	53.7	273	21	F03397
38	18.8	53.7	461	22	H07908
39	18.8	53.7	401	22	R44076
40	18.8	53.7	390	22	R52668
41	18.8	53.7	371	24	H84045
42	18.8	53.7	483	24	N33090
43	18.8	53.7	188	27	AA001067
44	18.8	53.7	613	27	AA001451
45	18.8	53.7	471	36	AA620375

ALIGNMENTS

RESULT 1
LOCUS AA190409/c
DEFINITION zp89c10.r1 Stratiogene Hela cell s3 937216 Homo sapiens cDNA clone IMAGE:627378 5' similar to SW-PIA-BOVIN P32871
ACCESSION AA190409
RNA sequence.
AA190409

NID 91779240
 VERSION AA190409.1 GI:1779240
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 417)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Donald, M.F., Chappel, B., Chissole, S., Dietrich, N., Dubaque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Maridis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rickin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E., Underwood, R., Woldmann, P., Waterston, R., Wilson, R. and Marra, M.
 TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478
 COMMENT On May 18, 1995 this sequence version replaced gi:811202.

CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LLM; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 328.
 Location/Qualifiers
 1..417
 /organism="Homo sapiens"
 /db_xref="GDB:5046076"
 /db_xref="taxon:9606"
 /clone="IMAGE:627378"
 /clone.lib="Stratagene HeLa cell s3 937216"
 /sex="female"
 /dev_stage="HeLa S3 cell line"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SK-; Site:1: EcoRI, Site:2: XhoI; Cloned unidirectionally. Primer: oligo dT. HeLa S3 epithelioid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb. Uni-ZAP XR vector. -5' adaptor sequence: 5' GATTCGGCCGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' "

BASE COUNT 116 a 98 c 99 t 11 others

Query Match 69.7%; Score 24.4; DB 29; Length 417;
 Best Local Similarity 96.2%; Pred. No. 1.9;
 Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 10 AGGCATCTAAAGCTACTATCATCC 35
 |||||
 Db 292 AGGCATCTTAAAGTCACTATCATCC 267

RESULT 2
 LOCUS AA610665/c
 DEFINITION np94803.s1 NCI-CGAP_Thy1 Homo sapiens cDNA clone IMAGE:1133932 similar to TR:G1236142 G1236142 MARINER TRANSPOSASE.; contains Alu repetitive element; contains element L1 repetitive element ;, mRNA sequence.
 ACCESSION AA610665 307 bp mRNA EST 09-DEC-1997
 NID 92459093
 VERSION AA610665.1 GI:2459093
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 307)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394357.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmerit-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Kitzman, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/dbp/image/image.html

Insert Length: 543 Std Error: 0.00
 Seq primer: -40M13 fwd: ET from Amersham
 High quality sequence stop: 297.
 Location/Qualifiers
 1..307
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1133932"
 /clone.lib="NCI-CGAP_Thy1"
 /tissue_type="thyroid"
 /lab_host="DH10B"
 /note="Vector: PAMP10; mRNA made from invasive thyroid tumor, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Kitzman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT 85 a 70 c 66 g 86 t

Query Match 59.4%; Score 20.8; DB 36; Length 307;
 Best Local Similarity 78.1%; Pred. No. 53;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 TTAAAGTTAGCATTCTTAAAGTCACTATCATC 33
 |||||
 Db 71 TAAAGTTTGGCTTCTCAAGACACTCATC 40

RESULT 3
 LOCUS AA887564/c
 DEFINITION oJ41912.s1 NCI-CGAP_k1d3 Homo sapiens cDNA clone IMAGE:1500934 3' similar to gb:M29870 RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 2 (HUMAN);, mRNA sequence.
 ACCESSION AA887564
 NID 93003252
 VERSION AA887564.1 GI:3003252
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 593)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2150626.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 DNA distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.llnl.gov/bdrip/image/image.html

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.llnl.gov/bdrip/image/image.html

FEATURES

SOURCE

Insert Length: 865 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 76.
 Location/Qualifiers

1. 593
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="700E03: 5: 219"
 /clone="IMAGE:1500934"
 /clone_id="NCI-CGAP_K1d3"
 /lab_host="DH10B"
 /note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 158 a 131 c 154 g 150 t
 ORIGIN

Query Match

Best Local Similarity 58.3%; Score 20.4; DB 39; Length 593;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TTAACTTAGGCATCTTAAGTCACATC 30
 ||||| ||||| ||||| ||||| |||||
 Db 290 TTAACTTAGGCATCTTAAGTCACATC 261

RESULT 4
 A1265938 389 bp mRNA EST 03-FEB-1999
 LOCUS q990f06.x1 Soares_total_fetus_NB2HF8_9w Homo sapiens CDNA clone
 DEFINITION IMAGE:1938659 3, mRNA sequence.
 ACCESSION A1265938
 MID g3884096
 VERSION A1265938.1 GI:3884096
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 389)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 17, 1998 this sequence version replaced gi:1900820.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550-

Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1499 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 384.
 Location/Qualifiers

FEATURES

SOURCE

1. 389
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1938659"
 /clone_id="Soares_total_fetus_NB2HF8_9w"
 /dev_stage="8-9 weeks"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGCGCGCTTAATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 121 a 71 c 87 g 110 t
 ORIGIN

Query Match

Best Local Similarity 58.3%; Score 20.4; DB 44; Length 389;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 AAGCTTAGGCATCTTAAGTCACATC 33
 ||||| ||||| ||||| ||||| |||||
 Db 352 AAGCTTCGTTTATCATCTACATCATC 381

RESULT 5
 A1543543 541 bp mRNA EST 22-MAR-1999
 LOCUS SD10336.Sprime SD Drosophila melanogaster Schneider L2 cell culture
 DEFINITION port2 Drosophila melanogaster CDNA clone SD10336 Sprime, mRNA
 sequence.
 ACCESSION A1543543
 MID g4460916
 VERSION A1543543.1 GI:4460916
 KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 541)
 Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brooks, P., Lewis, S. and Rubin, G.M.
 BDGP/HMT Drosophila EST Project
 JOURNAL Unpublished (1997)
 COMMENT On Mar 20, 1998 this sequence version replaced gi:2980225.

Contact: Harvey, D.
 G. M. Rubin-Molecular and Cell Biology
 University of California Berkeley
 539 LSA, Berkeley, CA 94720-3200, USA
 Fax: 510 643 9947.
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 Plate: 103 row: C column: 12
 High quality sequence stop: 488.
 Location/Qualifiers

FEATURES

SOURCE

1. 541
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /map="22: 19g13.30"
 /clone="SD10336"

/clone_lib="SD Drosophila melanogaster Schneider L2 cell culture port2"
/lab_host="DB5-alpha"
/note="Vector: port2; Site:1: EcoRI; Site:2: XhoI; Sized fractionated cDNAs were directly ligated into port2. Plasmid cDNA library."

BASE COUNT 176 a 105 c 118 g 142 t
ORIGIN

Query Match 57.7%; Score 20.2; DB 48; Length 541;
Best Local Similarity 75.8%; Pred. No. 90;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 TTAACTTAGGCACTCTAAAGTCACTATCATC 33
1 ||||| ||||| ||||| || |||||
Db 506 TATAGCTTAGGCTTTCTAAATCCCAATCAGC 474

RESULT 6 221049/c 314 bp MRNA EST 07-FEB-1995
LOCUS HSAAD42E H, Human adult Brain Cortex tissue Homo sapiens CDNA
DEFINITION similar to (Genbank M95258) human chromosome 4 sequence-tagged site
sts4-134, MRNA sequence.
221049 927792
VERSION 221049.1 GI:27792
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 314)
AUTHORS MRC Human Genome Mapping Project Resource Centre.
TITLE The UK-HGMP CDNA Program
JOURNAL Unpublished (1993)
COMMENT Contact: MRC Human Genome Mapping Project Resource Centre
Clinical Research Centre
Watford Road, Harrow, Middlesex HA1 3UD, U.K.
Email: biocore@hgmmp.mrc.ac.uk
*single read.

FEATURES
source Location/Qualifiers
1..314
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="H, Human adult Brain Cortex tissue"
/note="Vector: gtl1; clone library-H, Human adult Brain Cortex tissue; cloning vector is gtl1."
BASE COUNT 96 a 65 c 58 g 95 t
ORIGIN

Query Match 56.6%; Score 19.8; DB 20; Length 314;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 TTAACTTAGGCACTCTAAAGTCACTATCAT 32
1 ||||| ||||| ||||| || |||||
Db 138 TAAAGCTTAGGCTTTCTCAAAACACTTTTCA 108

RESULT 7
LOCUS A1161953/c 595 bp MRNA EST 03-DEC-1998
DEFINITION A010P20U Hybrid aspen plasmid library Populus tremula x Populus
tremuloides cDNA 5', MRNA sequence.
A1161953
ACCESSION A1161953
NID 93853238
VERSION A1161953.1 GI:3853238
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides.

ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliopsida; eudicotyledons; core
eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 595)
AUTHORS Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A.,
Holmberg, A., Amth, B., Bhalero, R., Larsson, M., Villarroel, R., Van
Montagu, M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W.,
Gustafsson, P., Uhlen, M., Sundberg, B., and Lundberg, J.
Gene discovery in the wood-forming tissues of poplar: Analysis of
5,692 expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
99007314
COMMENT On Jan 9, 1998 this sequence version replaced gi:937424.

TITLE
JOURNAL
MEDLINE

Contact: Sterky F
Department of Biotechnology
Royal Institute of Technology (KTH)
Teknikringen 34, S-100 44 STOCKHOLM, Sweden
Tel: +46 8 790 8287
Fax: +46 8 24 54 52
Email: fredrik@biochem.kth.se
PCR Primers
FORWARD: AAAGGGGATGCTGCAAGGCG
BACKWARD: GCTCCGCTGCTATCTTGTGCG
Seq primer: CATTCTAAAGCAGCGCCAG
High quality sequence stop: 595.
Location/Qualifiers
1..595
/organism="Populus tremula x Populus tremuloides"
/db_xref="taxon:47664"
/clone_lib="Hybrid aspen plasmid library"
/tissue_type="Cambial region"
/dev_stage="1.5 m actively growing tree"
/lab_host="E.coli"
/note="Vector: pBluescript SK; Site:1: SalI; Site:2: NotI;
Cambial region tissues, including developing xylem; the
meristematic cambial zone and the developing and mature
phloem, was harvested from 1.5 m actively growing trees.
cDNA was prepared and cloned into lambda gt22a. DNA was
isolated and subcloned into pBluescript SK using SalI and
NotI restriction enzymes."

BASE COUNT 169 a 120 c 139 g 163 t 4 others
ORIGIN

Query Match 56.6%; Score 19.8; DB 43; Length 595;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 3 TAAAGCTTAGGCACTCTAAAGTCACTATCATC 33
1 ||||| ||||| ||||| || |||||
Db 303 TAAAGCAAGCTTTAGAGTCACTACATC 273

RESULT 8
LOCUS A1162696/c 557 bp MRNA EST 03-DEC-1998
DEFINITION A022P09U Hybrid aspen plasmid library Populus tremula x Populus
tremuloides cDNA 5', MRNA sequence.
A1162696
ACCESSION A1162696
NID 93853981
VERSION A1162696.1 GI:3853981
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides.
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliopsida; eudicotyledons; core
eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 557)
AUTHORS Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A.,
Holmberg, A., Amth, B., Bhalero, R., Larsson, M., Villarroel, R., Van
Montagu, M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W.,

TITLE Gustafsson, P., Uhlen, M., Sundberg, B. and Lundberg, J.
Gene discovery in the wood-forming tissues of Poplar: Analysis of
5,692 expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
MEDLINE # 99007314
COMMENT On Sep 12, 1996 this sequence version replaced g1:1398120.

Contact: Sterky, F
Department of Biotechnology
Royal Institute of Technology (KTH)
Teknikringen 34, S-100 44 STOCKHOLM, Sweden
Tel: +46 8 790 8287
Fax: +46 8 24 54 52
Email: fredrik@biochem.kth.se
PCR Primers
FORWARD: AAAGGGGATGCTGCGACAGCG
BACKWARD: GCTCCGCTCGATGTGTGTC
Seq primer: CGTGTAAACGACGCGCAG
High quality sequence stop: 557.
Location/Qualifiers

FEATURES

source

1. 557
/organism="Populus tremula x Populus tremuloides"
/db_xref="taxon:47664"
/clone_lib="Hybrid aspen plasmid library"
/tissue_type="Cambial region"
/dev_stage="1.5 m actively growing tree"
/lab_host="E.coli"
/note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI;
Cambial region tissues, including developing xylem, the
meristematic cambial zone and the developing and mature
phloem, was harvested from 1.5 m actively growing trees.
cDNA was prepared and cloned into lambda gt22a. DNA was
isolated and subcloned into pBluescript SK using SalI and
NotI restriction enzymes."

BASE COUNT 164 a 107 c 125 g 156 t 5 others

ORIGIN

Query Match 56.6%; Score 19.8; DB 43; Length 557;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TAAGCTTAGCGATCTTAAGTCACATATCATC 33
||||| ||||| ||||| ||||| |||||
Db 309 TAAGCAAGCGCTTTTAAAGTCACATCATC 279

RESULT 9
LOCUS A1165066 491 bp mRNA EST 03-DEC-1998
DEFINITION A074p55u Hybrid aspen plasmid library Populus tremula x Populus
tremuloides cDNA 5', mRNA sequence.
ACCESSION A1165066
VERSION 93856351
KEYWORDS A1165066.1 GI:3856351
SOURCE EST.
ORGANISM Populus tremula x Populus tremuloides.
REFERENCE Populus tremula x Populus tremuloides.
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
euhylophytes: Spermatophyta: Magnoliophyta: eudicotyledons: core
eudicots: rosidae; eurosids I; Malpighiales; salicaceae; Populus.
1 (bases 1 to 491)
AUTHORS Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A.,
Holmberg, A., Amin, B., Bhalerao, R., Larsson, M., Villarroel, R., Van
Montagu, M., Sandberg, G., Olsson, O., Teerl, T.T., Boerjan, W.,
Gustafsson, P., Uhlen, M., Sundberg, B. and Lundberg, J.
TITLE Gene discovery in the wood-forming tissues of Poplar: Analysis of
5,692 expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
MEDLINE 99007314
COMMENT On Apr 18, 1995 this sequence version replaced g1:775315.
Contact: Sterky F

Department of Biotechnology
Royal Institute of Technology (KTH)
Teknikringen 34, S-100 44 STOCKHOLM, Sweden
Tel: +46 8 790 8287
Fax: +46 8 24 54 52
Email: fredrik@biochem.kth.se
PCR Primers
FORWARD: AAAGGGGATGCTGCGACAGCG
BACKWARD: GCTCCGCTCGATGTGTGTC
Seq primer: CGTGTAAACGACGCGCAG
High quality sequence stop: 491.
Location/Qualifiers

FEATURES

source

1. 491
/organism="Populus tremula x Populus tremuloides"
/db_xref="taxon:47664"
/map="7"
/clone_lib="Hybrid aspen plasmid library"
/tissue_type="Cambial region"
/dev_stage="1.5 m actively growing tree"
/lab_host="E.coli"
/note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI;
Cambial region tissues, including developing xylem, the
meristematic cambial zone and the developing and mature
phloem, was harvested from 1.5 m actively growing trees.
cDNA was prepared and cloned into lambda gt22a. DNA was
isolated and subcloned into pBluescript SK using SalI and
NotI restriction enzymes."

BASE COUNT 146 a 98 c 109 g 137 t 1 others

ORIGIN

Query Match 56.6%; Score 19.8; DB 43; Length 491;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TAAGCTTAGCGATCTTAAGTCACATATCATC 33
||||| ||||| ||||| ||||| |||||
Db 274 TAAGCAAGCGCTTTTAAAGTCACATCATC 244

RESULT 10
LOCUS A1188280 842 bp mRNA EST 13-OCT-1998
DEFINITION qd11h07.x1 Soares-placenta_8c9weeks_2MDHP8to9M Homo sapiens CDNA
clone IMAGE:1723453 3' similar to TR:Q13539 Q13539 MARINER
TRANSPOSASE. ;, mRNA sequence.
ACCESSION A1188280
VERSION 93739489
KEYWORDS A1188280.1 GI:3739489
SOURCE EST.
ORGANISM human.
REFERENCE Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;
Eutheria: Primates: Catarrhini: Hominiidae: Homo.
1 (bases 1 to 842)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Aug 21, 1998 this sequence version replaced.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from GIDCO
High quality sequence stop: 464.
Location/Qualifiers

1. 842
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1723453"

LOCUS A1314059 749 bp mRNA EST 17-DEC-1998
 DEFINITION U75B03.x1 Sugano mouse kidney mKia Mus musculus cDNA clone
 IMAGE:1920941 3' similar to SM:C10H_HUMAN Q16181 CDC10 PROTEIN
 HOMOLOG. ; mRNA sequence.
 ACCESSION A1314059
 MID 94029269
 VERSION A1314059.1 GI:4029269
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 749)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1798001.
 TITLE
 JOURNAL
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:977233
 Possible reversed clone: similarity on wrong strand
 Seq primer: custom primer used
 High quality sequence stop: 431.
 Location/Qualifiers
 1..749
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /map="21922"
 /clone="IMAGE:1920941"
 /clone_lib="Sugano mouse kidney mKia"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: Kidney; Vector: pME18S-FL3; Site1: DraIII
 (CACCTGTG); Site2: DraIII (CACCATGTG); 1st strand cDNA
 was primed with an oligo(dT) primer
 [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was
 ligated to a DraIII adaptor [TGTGGCCCTGTG], digested
 and cloned into distinct DraIII sites of the pME18S-FL3
 vector (5' site CACCTGTG, 3' site CACCATGTG). XhoI should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments <1.5kb. Library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for
 sequencing: 5' end primer CTCTGCTCTTAAGAGTGGC and 3' end
 primer CGACCTGACGCGAGCA."
 BASE COUNT 160 a 179 c 112 g 294 t 4 others
 ORIGIN
 Query Match 56.0%; Score 19.6; DB 44; Length 749;
 Best Local Similarity 73.5%; Pred. No. 1.5e+02;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 Oy 2 TTAAGCTTAGGCATTTCTAAGTCATCATCCTCC 35
 Db 680 TGAACGTTGGCATGAAGAAAGTAATATCATCCTCC 647
 RESULT :14

HSN002526
 ID HSN002526 standard; RNA; EST; 620 bp.
 AC AL038186;
 XX AL038186.1
 XX SV e1396521
 NI e1396521
 XX
 DT 12-MAR-1999 (Rel. 59, Created)
 DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
 XX Homo sapiens mRNA; EST DKFZP566N1624_r1 (from clone
 DE DKFZP566N1624)
 DE EST; expressed sequence tag.
 XX
 OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
 OC Primates; Catarrhini; Homidae; Homo.
 XX
 [1]
 RA 1-620
 RA Blocker H., Boecher M., Brandt P., Mewes W., Gassenhuber J.,
 Wiemann S.;
 RT Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
 RL MIPS, Am Kioferspitze 18a D-82152 Martinsried, GERMANY
 XX
 CC Clone from S. Wiemann, sequenced by GBF within the cDNA
 CC sequencing consortium of the German Genome Project
 CC s1 sequence also available
 CC This clone is available at the RZPD in Berlin
 CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
 XX
 XX Key Location/Qualifiers
 FH
 FT source 1..620
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT /clone="DKFZP566N1624"
 FT /clone_lib="566 (synonym: hKfd2). Vector pAMP1; host
 FT Xl-2Blue; sites NotI + SalI"
 FT /dev_stage="fetal"
 FT /tissue_type="kidney"
 FT
 XX
 SO Sequence 620 bp; 196 A; 116 C; 145 G; 163 T; 0 other;
 Query Match 56.0%; Score 19.6; DB 53; Length 620;
 Best Local Similarity 73.5%; Pred. No. 1.5e+02;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 Oy 2 TTAAGCTTAGGCATTTCTAAGTCATCATCCTCC 35
 Db 533 TGAACGTTGGCATGAAGAAAGTAATATCATCCTCC 566
 RESULT :15
 HSN002976
 ID HSN002976 standard; RNA; EST; 762 bp.
 AC AL038500;
 XX AL038500.1
 XX SV e1396835
 NI e1396835
 XX
 DT 12-MAR-1999 (Rel. 59, Created)
 DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
 XX Homo sapiens mRNA; EST DKFZP566D1246_r1 (from clone
 DE

```

DE DKEZP566D1246)
XX
XX EST: expressed sequence tag.
XX
XX Homo sapiens (human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC Primates; Catarrhini; Homnidae; Homo.
XX
XX [1]
RN 1-762
RP Ottenwaelder B., Obermaier B., Mewes W., Gassenhuber J., Wiemann S.;
RT Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
RL MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
XX
XX Clone from S. Wiemann, sequenced by Medigenomix within the CDNA
CC sequencing consortium of the German Genome Project
CC s1 sequence also available
CC This clone is available at the RZPD in Berlin
CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
XX Key location/Qualifiers
FH 1. 762
FT source
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="DKEZP566D1246"
FT /clone_lib="566 (synonym: hfk42). Vector pAMP1; host
FT X1-2blue; sites NotI + SalI"
FT /dev_stage="fetal"
FT /tissue_type="kidney"
XX
SQ Sequence 762 bp; 262 A; 125 C; 164 G; 211 T; 0 other;

```

```

Query Match 56.0%; Score 19.6; DB 53; Length 762;
Best Local Similarity 73.5%; Pred. No. 1.5e+02;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 2 TTAAGCTTAGCATCTTAAGTCAATATCATCCC 35
| |||| | |||| | |||| | |||| |
DB 396 TGAAGCCTTGCAATGAAAAGTGAATATCATCCC 429

```

Search completed: September 13, 1999, 15:45:56
 Job time: 4044 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:23 ; Search time 539.84 Seconds

(without alignments)
106.042 Million cell updates/sec

Title: US-09-325-095-18

Perfect score: 18
Sequence: 1 GACTCGAGTCGACATCGA 18

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database:

GenBankl:*

1: gb_ba1:*

2: gb_ba2:*

3: gb_om:*

4: gb_ov:*

5: gb_pat:*

6: gb_ph:*

7: gb_pl1:*

8: gb_pl2:*

9: gb_pl1:*

10: gb_pl2:*

11: gb_pl3:*

12: gb_pl4:*

13: gb_pl5:*

14: gb_pl6:*

15: gb_pl7:*

16: gb_pl8:*

17: gb_pl9:*

18: gb_pl10:*

19: gb_pl11:*

20: gb_pl12:*

21: gb_pl13:*

22: gb_pl14:*

23: gb_pl15:*

24: gb_pl16:*

25: gb_pl17:*

26: gb_pl18:*

27: gb_pl19:*

28: gb_pl20:*

29: gb_pl21:*

30: gb_pl22:*

31: gb_pl23:*

32: gb_pl24:*

33: gb_pl25:*

34: gb_pl26:*

35: gb_pl27:*

36: gb_pl28:*

37: gb_pl29:*

38: gb_pl30:*

39: gb_pl31:*

40: gb_pl32:*

41: gb_pl33:*

42: gb_pl34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	ID	Description
1	18	100.0	638 4	XU088561
				088561 Xenopus lae

Result No.	Score	Match Length	ID	Description
2	18	100.0	35 5	A24382
3	18	100.0	35 5	A27645
4	18	100.0	35 5	A37244
5	18	100.0	35 5	A37246
6	18	100.0	35 5	A40372
7	18	100.0	35 5	A40373
8	18	100.0	35 5	A42335
9	18	100.0	35 5	A42384
10	18	100.0	35 5	A46467
11	18	100.0	35 5	A59198
12	18	100.0	20 5	A59199
13	18	100.0	35 5	A58608
14	18	100.0	35 5	AR012367
15	18	100.0	35 5	AR012367
16	18	100.0	35 5	AR012367
17	18	100.0	35 5	AR012367
18	18	100.0	35 5	AR012367
19	18	100.0	35 5	AR012367
20	18	100.0	35 5	AR012367
21	18	100.0	35 5	AR012367
22	18	100.0	35 5	AR012367
23	18	100.0	35 5	AR012367
24	18	100.0	35 5	AR012367
25	18	100.0	35 5	AR012367
26	18	100.0	35 5	AR012367
27	18	100.0	35 5	AR012367
28	18	100.0	35 5	AR012367
29	18	100.0	35 5	AR012367
30	18	100.0	35 5	AR012367
31	18	100.0	35 5	AR012367
32	18	100.0	35 5	AR012367
33	18	100.0	35 5	AR012367
34	18	100.0	35 5	AR012367
35	18	100.0	35 5	AR012367
36	18	100.0	35 5	AR012367
37	18	100.0	35 5	AR012367
38	18	100.0	35 5	AR012367
39	18	100.0	35 5	AR012367
40	18	100.0	35 5	AR012367
41	18	100.0	35 5	AR012367
42	18	100.0	35 5	AR012367
43	18	100.0	35 5	AR012367
44	18	100.0	35 5	AR012367
45	18	100.0	35 5	AR012367

ALIGNMENTS

Result No.	Score	Match Length	ID	Description
1	18	100.0	638 bp	XU088561 mRNA
				Xenopus laevis E2 ubiquitin conjugating enzyme (Ubc9) mRNA,
				complete cds.
				Accession
				U88561
				Version
				92078330
				Keywords
				U88561.1 GI:2078330
				Source
				African clawed frog.
				Organism
				Xenopus laevis
				Eukaryota; Metazoa; Chordata; Vertebrata; Amphibia;
				Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae;
				Xenopus.
				1 (bases 1 to 638)
				Saitoh, H., Pu, R., Cavenagh, M. and Dasso, M.
				RanBP2 associates with Ubc9p and a modified form of RanGAP1
				Proc. Natl. Acad. Sci. U.S.A. 94 (8), 3736-3741 (1997)
				2 (bases 1 to 638)
				Saitoh, H., Pu, R., Cavenagh, M. and Dasso, M.
				Direct Submission
				Submitted (05-FEB-1997) LME/NICHD, NIH, Bldg. 18, Rm. 101, Bethesda,
				MD 20892, USA

FEATURES
source location/Qualifiers
1. .638
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/cell_type="oocyte"
107. .583
/gene="ubc9"
/gene="ubc9"
/note="ubc9"
/codon_start=1
/product="E2 ubiquitin conjugating enzyme"
/protein_id="AA57336.1"
/db_xref="PIR:P07833.1"
/translation="MSGIALSLAQRKAKRDPGFVAVPTKNDGTMNLMNECA
IFGKKGTPWEGGLFRLMLFKDDYPSPPKCPLEFPHNYPGSGTGLSIEEDKD
WRPATIKILIGIOELNBPNIODPAQAEAVTIYQNVETKRRVRAQAKKFAFS"

BASE COUNT 214 a 131 c 147 g 146 t
ORIGIN

Query Match 100.0%; Score 18; DB 4; Length 638;
Best Local Similarity 100.0%; Pred. No. 6.1; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||

DB 638 GACTCGAGTCGACATCGA 621

RESULT 2
A24382 35 bp DNA PAT 21-MAR-1995
LOCUS A24382 35 bp DNA
DEFINITION d117 adaptor primer.
ACCESSION A24382
NID 9904477
VERSION A24382.1 GI:904477
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 35)
AUTHORS Samlencos, P., De Taxis du Poet, P., Mittl, G. and Scaccheri, E.
TITLE Anti-thrombin polypeptides
JOURNAL Patent: EP 0501821-A 22 02-SEP-1992;
FARMITALIA CARLO ERBA S.r.L.
FEATURES location/Qualifiers
source 1. .35
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||

DB 1 GACTCGAGTCGACATCGA 18

RESULT 3
A27645 35 bp DNA PAT 04-JUN-1995
LOCUS A27645 35 bp DNA
DEFINITION Hybrid d117-adaptor primer.
ACCESSION A27645
NID 91248481
VERSION A27645.1 GI:1248481
KEYWORDS
SOURCE synthetic construct.

ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 35)
AUTHORS POLYPHENOL OXIDASE GENES
JOURNAL Patent: WO 9302195-A 1 04-FEB-1993;
FEATURES location/Qualifiers
source 1. .35
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||

DB 1 GACTCGAGTCGACATCGA 18

RESULT 4
A37244 35 bp DNA PAT 05-MAR-1997
LOCUS A37244 35 bp DNA
DEFINITION Sequence 13 from Patent WO9403609.
ACCESSION A37244
NID 92294355
VERSION A37244.1 GI:2294355
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
TITLE EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 13 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES location/Qualifiers
source 1. .35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||

DB 1 GACTCGAGTCGACATCGA 18

RESULT 5
A37246 18 bp DNA PAT 05-MAR-1997
LOCUS A37246 18 bp DNA
DEFINITION Sequence 15 from Patent WO9403609.
ACCESSION A37246
NID 92294357
VERSION A37246.1 GI:2294357
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 18)
AUTHORS Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
TITLE EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM

JOURNAL Patent: WO 9403609-A 15 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES Location/Qualifiers
Source 1.18
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 3 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 6
LOCUS A40372 18 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 28 from Patent WO9425489.
ACCESSION A40372
NID 92296421
VERSION A40372.1 GI:2296421
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 18)
AUTHORS Mohapatra,S.S. and Sehon,A.H.
TITLE DOWNREGULATION OF ALLERGEN-SPECIFIC IMMUNE RESPONSE
JOURNAL UNIV MANITOBA (CA)
Patent: WO 9425489-A 28 10-NOV-1994;
COMMENT Other publication AU 6674094 941121.
FEATURES Location/Qualifiers
Source 1.18
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 3 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 7
LOCUS A40373 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 29 from Patent WO9425489.
ACCESSION A40373
NID 92296422
VERSION A40373.1 GI:2296422
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Mohapatra,S.S. and Sehon,A.H.
TITLE DOWNREGULATION OF ALLERGEN-SPECIFIC IMMUNE RESPONSE
JOURNAL UNIV MANITOBA (CA)
Patent: WO 9425489-A 29 10-NOV-1994;
COMMENT Other publication AU 6674094 941121.
FEATURES Location/Qualifiers
Source 1.35

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 8
LOCUS A42335 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 7 from Patent WO9502057.
ACCESSION A42335
NID 92297812
VERSION A42335.1 GI:2297812
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Gusterson,B.A., Crompton,M.R., Mitchell,P.J., Barker,K.T.,
Kamali,T., Page,M.J. and Spence,P.
TITLE PROTEIN TYROSINE KINASE AND LIGANDS THEREOF
JOURNAL CANCER RES INST (GB)
Patent: WO 9502057-A 7 19-JAN-1995;
COMMENT Other publication AU 7080994 950206.
FEATURES Location/Qualifiers
Source 1.35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 9
LOCUS A42384 35 bp DNA PAT 06-MAR-1997
DEFINITION Sequence 7 from Patent WO9502187.
ACCESSION A42384
NID 92297858
VERSION A42384.1 GI:2297858
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Gusterson,B.A., Crompton,M.R., Mitchell,P.J., Barker,K.T.,
Martindale,J.E., Page,M.J. and Spence,P.
TITLE CELL GROWTH FACTOR RECEPTORS
JOURNAL CANCER RES INST (GB)
Patent: WO 9502187-A 7 19-JAN-1995;
COMMENT Other publication AU 7081094 950206.
FEATURES Location/Qualifiers
Source 1.35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 10

LOCUS A46467 35 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 5 from Patent WO9526402.
ACCESSION A46467
NID 92300644

VERSION A46467.1 GI:2300644

KEYWORDS

SOURCE

REFERENCE 1 (bases 1 to 35)
AUTHORS Knox,D.P., Smith,S.K., Smith,W.D., Redmond,D. and Murray,J.
TITLE VACCINES AGAINST HELMINTHIC PARASITES
JOURNAL MALDIKRODT VETERINARY INC (US)
Patent: WO 9526402-A 5 05-OCT-1995;

COMMENT Other Publication ZA 9502370 951215
Other Publication CA 2182178 951005
Other Publication AU 1956495 951017.

FEATURES

source
1..35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 11

LOCUS A59198 40 bp DNA PAT 06-MAR-1998
DEFINITION Sequence 7 from Patent WO9704108.
ACCESSION A59198
NID 93714579

VERSION A59198.1 GI:3714579

KEYWORDS

SOURCE

REFERENCE 1 (bases 1 to 40)
AUTHORS Schuster,E., Spreesler,B., Tltze,K., Gottschalk,M., Khanh,N.Q.,
TITLE WOLF,S. and Plainer,H.
JOURNAL LECITINE AMINOPEPTIDASES PRODUCED RECOMBINANTLY FROM ASPERGILLUS
Patent: WO 9704108-A 7 06-FEB-1997;
ROEHM GMBH (DE)
Other Publication DE 19526485 970123.

COMMENT

FEATURES

source
1..40
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 24 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 40;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 12

LOCUS A59199 20 bp DNA PAT 06-MAR-1998
DEFINITION Sequence 8 from Patent WO9704108.
ACCESSION A59199
NID 93714580

VERSION A59199.1 GI:3714580

KEYWORDS

SOURCE

REFERENCE 1 (bases 1 to 20)
AUTHORS Schuster,E., Spreesler,B., Tltze,K., Gottschalk,M., Khanh,N.Q.,
TITLE WOLF,S. and Plainer,H.
JOURNAL LECITINE AMINOPEPTIDASES PRODUCED RECOMBINANTLY FROM ASPERGILLUS
Patent: WO 9704108-A 8 06-FEB-1997;
ROEHM GMBH (DE)
Other Publication DE 19526485 970123.

COMMENT

FEATURES

source
1..20
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 5 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 13

LOCUS A68608 35 bp DNA PAT 06-MAY-1999
DEFINITION Sequence 8 from Patent WO9749726.
ACCESSION A68608
NID 94759636

VERSION A68608.1 GI:4759636

KEYWORDS

SOURCE

REFERENCE 1 (bases 1 to 35)
AUTHORS Mele,A., De,S.R., Parente,D. and Colnaghi,M.I.
TITLE RECOMBINANT RIBOSOMAL INHIBITOR PROTEIN (RIP) AND USE AS
JOURNAL IMMUNOCONJUGATE
Patent: WO 9749726-A 31-DEC-1997;
MINISTERO UNI RICERCA SCIENT E (IT)
Other Publication IT FI960155 19971229.

COMMENT

FEATURES

source
1..35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Job time: 4551 sec

Query Match 100.0%; Score 18; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 |||||||
 Db 1 GACTCGAGTCGACATCGA 18

RESULT 14

ARO12367 ARO12367 35 bp DNA PAT 04-DEC-1998
 LOCUS Sequence 5 from patent US 5763400.
 DEFINITION ARO12367
 ACCESSION ARO12367
 NID 93970357
 VERSION ARO12367.1 GI:3970357
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 35)
 AUTHORS Adams, M.E. and Altman, D.
 TITLE Ecdysis-triggering hormone compositions
 JOURNAL Patent: US 5763400-A 5 09-JUN-1998;
 FEATURES Location/Qualifiers
 source 1..35

BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN /organism="unknown"

Query Match 100.0%; Score 18; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 |||||||
 Db 1 GACTCGAGTCGACATCGA 18

RESULT 15

I13679 I13679 35 bp DNA PAT 08-AUG-1995
 LOCUS Sequence 35 from patent US 5439820.
 DEFINITION I13679
 ACCESSION I13679
 NID 9996745
 VERSION I13679.1 GI:996745
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 35)
 AUTHORS Sarmientos, P., De Taxis du Poet, P., Nitli, G. and Scacheri, E.
 TITLE Anti-thrombin polypeptides
 JOURNAL Patent: US 5439820-A 35 08-AUG-1995;
 FEATURES Location/Qualifiers
 source 1..35

BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN /organism="unknown"

Query Match 100.0%; Score 18; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 |||||||
 Db 1 GACTCGAGTCGACATCGA 18

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

UM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:31 ; Search time 255.05 Seconds

(without alignments)
17.657 Million cell updates/sec

Title: US-09-325-095-18

Perfect score: 18

Sequence: 1 GACTCGAGTCGACATCGA 18

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N.Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	39	1	Q23200
2	18	100.0	32	1	Q23201
3	18	100.0	22	1	Q23202
4	18	100.0	38	1	Q23203
5	18	100.0	39	1	Q23205
6	18	100.0	39	1	Q28423
7	18	100.0	32	1	Q28424
8	18	100.0	22	1	Q28425
9	18	100.0	38	1	Q31777
10	18	100.0	32	1	Q31778
11	18	100.0	35	1	Q36668
12	18	100.0	18	1	Q48416
13	18	100.0	39	1	Q52547
14	18	100.0	35	1	Q52503
15	18	100.0	35	1	Q59015
16	18	100.0	18	1	Q59017
17	18	100.0	35	1	Q57021
18	18	100.0	18	1	Q57023
19	18	100.0	19	1	Q44401
20	18	100.0	35	1	Q71021
21	18	100.0	35	1	Q63873
22	18	100.0	35	1	Q65594
23	18	100.0	18	1	Q78806
24	18	100.0	35	1	Q78807
25	18	100.0	35	1	Q84786
26	18	100.0	32	1	Q86055
27	18	100.0	38	1	Q86054
28	18	100.0	35	1	Q89167
29	18	100.0	35	1	Q89878
30	18	100.0	35	1	Q89878
31	18	100.0	1296	1	T05321
32	18	100.0	35	1	T05325
33	18	100.0	35	1	Q94244
34	18	100.0	19	1	T10697
35	18	100.0	901	1	T10698
36	18	100.0	35	1	T16211
37	18	100.0	35	1	T10276
38	18	100.0	35	1	T03701
39	18	100.0	35	1	T11753
40	18	100.0	18	1	T11754
41	18	100.0	37	1	Q74002
42	18	100.0	27	1	Q74003
43	18	100.0	35	1	T35802

44 18 100.0 35 1 T28797
45 18 100.0 18 1 V56009

MMIV reverse trans
D. discoideum Dp1

ALIGNMENTS

RESULT 1
Q23200 Q23200 standard; DNA; 39 BP.
AC Q23200:
DE 29-JUL-1992 (first entry)
DE ZC2487 oligo-d(T), XhoI, SalI, ClaI primer for AOH DNA.
KW Acyl:oxyacyl hydrolase; septicemia; LPS; vaccine; human; ss.
OS Synthetic.
PN WO9204444-A.
PD 19-MAR-1992.
PF 11-SEP-1991; U06569.
PR 12-SEP-1990; US-581342.
PA (ZYMO-) ZYMOGENETICS INC.
PI (TEXA) UNIV OF TEXAS SYSTEM.
DR Ohara PJ, Hagen FS, Grant FJ, Munford RS;
PI WPI; 92-114352/14.
PT Prodn. of acyl:oxyacyl-hydrolase in eukaryotes - using DNA
PI constructed encoding whole protein or its sub-units, and used to
PI treat gram negative sepsis.
PS Table 1; Page 29; 97pp; English.
CC The primer was used to synthesize cDNA from U-937 polyA+ RNA.
CC Both 5' and 3' templates were prepd. The resulting cDNAs were
CC enriched and amplified, and were used for the isolation of the
CC complete gene sequence for AOH. The gene was used to express
CC the recombinant enzyme, which can be used prophylactically and
CC therapeutically against Gram-negative septicemia in mammals.
CC For treatment, 1 ug-10 mg AOH is given per 70 kg body wt. The
CC recombinant enzyme is produced in higher levels than found
CC naturally in neutrophils and is more easily purified. See also
CC Q23190-215.
SQ Sequence 39 BP: 6 A; 6 C; 6 G; 21 T;

Query Match 100.0%; Score 18; DB 1; Length 39;
Best local similarity 100.0%; Pred. No. 0.33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
DB 1 GACTCGAGTCGACATCGA 18

RESULT 2
Q23201 Q23201 standard; DNA; 32 BP.
AC Q23201:
DE 29-JUL-1992 (first entry)
DE ZC2488 oligo-d(C), XhoI, SalI, ClaI primer for AOH DNA.
KW Acyl:oxyacyl hydrolase; septicemia; LPS; vaccine; human; ss.
OS Synthetic.
PN WO9204444-A.
PD 19-MAR-1992.
PF 11-SEP-1991; U06569.
PR 12-SEP-1990; US-581342.
PA (ZYMO-) ZYMOGENETICS INC.
PI (TEXA) UNIV OF TEXAS SYSTEM.
DR Ohara PJ, Hagen FS, Grant FJ, Munford RS;
PI WPI; 92-114352/14.
PT Prodn. of acyl:oxyacyl-hydrolase in eukaryotes - using DNA
PI constructed encoding whole protein or its sub-units, and used to
PI treat gram negative sepsis.
PS Table 1; Page 29; 97pp; English.
CC The primer was used for second strand synthesis of cDNA from U-937
CC polyA+ RNA. Both 5' and 3' templates were prepd. The resulting
CC cDNAs were enriched and amplified, and were used for the isolation
CC of the complete gene sequence for AOH. The gene was used to

CC express the recombinant enzyme, which can be used prophylactically
 CC and therapeutically against Gram-negative septicemia in mammals.
 CC For treatment, 1 ug-10 mg AOH is given per 70 kg body wt. The
 CC recombinant enzyme is produced in higher levels than found
 CC naturally in neutrophils and is more easily purified. See also
 CC Q23190-215.
 SQ Sequence 32 BP; 6 A; 16 C; 6 G; 4 T;

Query Match 100.0%; Score 18; DB 1; Length 32;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 1 GACTCGAGTCGACATCGA 18

RESULT 3
 ID Q23202 standard; DNA: 22 BP.
 AC Q23202;
 DT 29-JUL-1992 (first entry)
 DE ZC2489 primer for AOH DNA.
 KW Acyl oxyacyl hydrolase; septicemia; LPS; vaccine; human; ss.
 OS Synthetic.
 PN WO9204444-A.
 PD 19-MAR-1992.
 PF 11-SEP-1991; U06569.
 PR 12-SEP-1990; US-581342.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI (TEXA) UNIV OF TEXAS SYSTEM.
 PI Ohara PJ, Hagen FS, Grant FJ, Munford RS;
 DR WPI: 92-114352/14.
 PT Prodn. of acyl:oxy:acyl-hydrolase in eukaryotes - using DNA
 PT constructed encoding whole protein or its sub-units, and used to
 PT treat gram negative sepsis.
 PS Table 1; Page 29; 97pp; English.
 CC The primer was used for PCR amplification of cDNA from U-937
 CC polyA+ RNA. The resulting DNA was used for the isolation
 CC of the complete gene sequence for AOH. The gene was used to
 CC express the recombinant enzyme, which can be used prophylactically
 CC and therapeutically against Gram-negative septicemia in mammals.
 CC For treatment, 1 ug-10 mg AOH is given per 70 kg body wt. The
 CC recombinant enzyme is produced in higher levels than found
 CC naturally in neutrophils and is more easily purified. See also
 CC Q23190-215.
 SQ Sequence 22 BP; 6 A; 6 C; 6 G; 4 T;

Query Match 100.0%; Score 18; DB 1; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.32;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 1 GACTCGAGTCGACATCGA 18

RESULT 4
 ID Q23203 standard; DNA: 38 BP.
 AC Q23203;
 DT 29-JUL-1992 (first entry)
 DE ZC2633 primer for AOH DNA.
 KW Acyl oxyacyl hydrolase; septicemia; LPS; vaccine; human; ss.
 OS Synthetic.
 PN WO9204444-A.
 PD 19-MAR-1992.
 PF 11-SEP-1991; U06569.
 PR 12-SEP-1990; US-581342.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI (TEXA) UNIV OF TEXAS SYSTEM.

PI Ohara PJ, Hagen FS, Grant FJ, Munford RS;
 DR WPI: 92-114352/14.
 PT Prodn. of acyl:oxy:acyl-hydrolase in eukaryotes - using DNA
 PT constructed encoding whole protein or its sub-units, and used to
 PT treat gram negative sepsis.
 PS Table 1; Page 29; 97pp; English.
 CC The primer was used for PCR amplification of cDNA from U-937
 CC polyA+ RNA. The resulting DNA was used for the isolation
 CC of the complete gene sequence for AOH. The gene was used to
 CC express the recombinant enzyme, which can be used prophylactically
 CC and therapeutically against Gram-negative septicemia in mammals.
 CC For treatment, 1 ug-10 mg AOH is given per 70 kg body wt. The
 CC recombinant enzyme is produced in higher levels than found
 CC naturally in neutrophils and is more easily purified. See also
 CC Q23190-215.
 SQ Sequence 38 BP; 11 A; 9 C; 12 G; 6 T;

Query Match 100.0%; Score 18; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 17 GACTCGAGTCGACATCGA 34

RESULT 5
 ID Q23205 standard; DNA: 39 BP.
 AC Q23205;
 DT 29-JUL-1992 (first entry)
 DE ZC2632 primer for AOH DNA.
 KW Acyl oxyacyl hydrolase; septicemia; LPS; vaccine; human; ss.
 OS Synthetic.
 PN WO9204444-A.
 PD 19-MAR-1992.
 PF 11-SEP-1991; U06569.
 PR 12-SEP-1990; US-581342.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI (TEXA) UNIV OF TEXAS SYSTEM.
 PI Ohara PJ, Hagen FS, Grant FJ, Munford RS;
 DR WPI: 92-114352/14.
 PT Prodn. of acyl:oxy:acyl-hydrolase in eukaryotes - using DNA
 PT constructed encoding whole protein or its sub-units, and used to
 PT treat gram negative sepsis.
 PS Table 1; Page 29; 97pp; English.
 CC The primer was used for PCR amplification of cDNA from U-937
 CC polyA+ RNA. The resulting DNA was used for the isolation
 CC of the complete gene sequence for AOH. The gene was used to
 CC express the recombinant enzyme, which can be used prophylactically
 CC and therapeutically against Gram-negative septicemia in mammals.
 CC For treatment, 1 ug-10 mg AOH is given per 70 kg body wt. The
 CC recombinant enzyme is produced in higher levels than found
 CC naturally in neutrophils and is more easily purified. See also
 CC Q23190-215.
 SQ Sequence 39 BP; 13 A; 10 C; 10 G; 6 T;

Query Match 100.0%; Score 18; DB 1; Length 39;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 18 GACTCGAGTCGACATCGA 35

RESULT 6
 ID Q28423 standard; cDNA: 39 BP.
 AC Q28423;
 DT 25-FEB-1993 (first entry)

DE Human galanin cDNA PCR primer ZC2487.
 KW Polymerase chain reaction; ss.
 OS Synthetic.
 PN WO9215015-A.
 PD 03-SEP-1992.
 PF 25-FEB-1992; U01469.
 PR 25-FEB-1991; US-662221.
 PS 31-DEC-1991; US-816285.
 PA (UYBO-) UNIV BOARDS WASHINGTON.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI Kowalyk S, McKnight GL, Smith RA;
 DR WPI; 92-316309/38.
 PT Detecting presence of galanin antagonist - by exposing isolated
 PT galanin analogue to galanin receptor in presence of natural
 PS Example; Page 25; 73pp; English.
 CC The sequence is that of a degenerate primer encoding a galanin DNA
 CC sequence which was used to amplify galanin cDNA sequences by the
 CC polymerase chain reaction (PCR) method. It also contains a 5' tail
 CC of 10 nucleotides for subcloning. See also Q28421-Q28432.
 SQ Sequence 39 BP; 6 A; 6 C; 6 G; 21 T;

Query Match 100.0%; Score 18; DB 1; Length 39;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 |||||
 DB 1 GACTCGAGTCGACATCGA 18

RESULT 7
 Q28424
 ID Q28424 standard; cDNA; 32 BP.
 AC Q28424;
 DT 25-FEB-1993 (first entry)
 DE Human galanin cDNA PCR primer ZC2488.
 KW Polymerase chain reaction; ss.
 OS Synthetic.
 PN WO9215015-A.
 PD 03-SEP-1992.
 PF 25-FEB-1992; U01469.
 PR 25-FEB-1991; US-662221.
 PS 31-DEC-1991; US-816285.
 PA (UYBO-) UNIV BOARDS WASHINGTON.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI Kowalyk S, McKnight GL, Smith RA;
 DR WPI; 92-316309/38.
 PT Detecting presence of galanin antagonist - by exposing isolated
 PT galanin analogue to galanin receptor in presence of natural
 PS Example; Page 25; 73pp; English.
 CC The sequence is that of a degenerate primer encoding a galanin DNA
 CC sequence which was used to amplify galanin cDNA sequences by the
 CC polymerase chain reaction (PCR) method. It also contains a 5' tail
 CC of 10 nucleotides for subcloning. See also Q28421-Q28432.
 SQ Sequence 32 BP; 6 A; 16 C; 6 G; 4 T;

Query Match 100.0%; Score 18; DB 1; Length 32;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 |||||
 DB 1 GACTCGAGTCGACATCGA 18

RESULT 8
 Q28425
 ID Q28425 standard; cDNA; 22 BP.
 AC Q28425;

DT 25-FEB-1993 (first entry)
 DE Human galanin cDNA PCR primer ZC2489.
 KW Polymerase chain reaction; ss.
 OS Synthetic.
 PN WO9215015-A.
 PD 03-SEP-1992.
 PF 25-FEB-1992; U01469.
 PR 25-FEB-1991; US-662221.
 PS 31-DEC-1991; US-816285.
 PA (UYBO-) UNIV BOARDS WASHINGTON.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI Kowalyk S, McKnight GL, Smith RA;
 DR WPI; 92-316309/38.
 PT Detecting presence of galanin antagonist - by exposing isolated
 PT galanin analogue to galanin receptor in presence of natural
 PS Example; Page 25; 73pp; English.
 CC The sequence is that of a degenerate primer encoding a galanin DNA
 CC sequence which was used to amplify galanin cDNA sequences by the
 CC polymerase chain reaction (PCR) method. It also contains a 5' tail
 CC of 10 nucleotides for subcloning. See also Q28421-Q28432.
 SQ Sequence 22 BP; 6 A; 6 C; 6 G; 4 T;

Query Match 100.0%; Score 18; DB 1; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.32;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 |||||
 DB 1 GACTCGAGTCGACATCGA 18

RESULT 9
 Q31777
 ID Q31777 standard; DNA; 38 BP.
 AC Q31777;
 DT 20-APR-1993 (first entry)
 DE Human pancreatic islet cell GAD PCR primer ZC2633.
 KW Human pancreatic islet cell GAD PCR primer ZC2633.
 OS Synthetic.
 PN WO9220811-A.
 PD 26-NOV-1992.
 PF 14-MAY-1992; U04079.
 PR 15-MAY-1991; US-702162.
 PA (UNIV) UNIV WASHINGTON.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI Foster DC, Grublin CE, Hagopian W, Karlsten AE, Lermmark A, Ohara PJ;
 DR WPI; 92-415785/50.
 PT Polynucleotide encoding human islet glutamic acid decarboxylase -
 PT used to test for auto-antibodies against itself to diagnose
 PT insulin dependent diabetes mellitus
 PS Example; Page 25; 45pp; English.
 CC The sequence is that of an oligonucleotide primer ZC2633 which
 CC is complementary to the EcoRI adapter and was used in the second
 CC PCR step to prime the plus strand. It was used to obtain full
 CC length human pancreatic islet GAD cDNA clones.
 SQ Sequence 38 BP; 11 A; 9 C; 12 G; 6 T;

Query Match 100.0%; Score 18; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 |||||
 DB 17 GACTCGAGTCGACATCGA 34

RESULT 10
 Q31778
 ID Q31778 standard; DNA; 32 BP.
 AC Q31778;

DT 20-APR-1993 (first entry)
 DE Human pancreatic islet cell GAD PCR primer ZC2488.
 KW Polymerase chain reaction; glutamic acid decarboxylase; ss.
 OS Synthetic.
 PN WO9220811-A.
 PD 26-NOV-1992.
 PF 14-MAY-1992; US-702162.
 PR 15-MAY-1991; US-702162.
 PA (UNIW) UNIV WASHINGTON.
 PI (ZYMO) ZYMOGENETICS INC.
 DR Foster DC, Grublin CE, Hagopian W, Karlson AE, Lernmark A, Ohara RJ,
 WPI: 93-415789/50.
 PT Polynucleotide encoding human islet glutamic acid decarboxylase -
 used to test for auto-antibodies against itself to diagnose
 PT Insulin dependent diabetes mellitus
 PS Example: Page 25; 45pp; English.
 CC The sequence is that of an oligonucleotide primer ZC2488 with an
 CC EcoRI adapter which was used on second strand synthesis to
 CC generate a cDNA population enriched for glutamic acid decarboxylase
 CC (GAD) but still heterogeneous due to non-specific pairing of the
 CC internal primer. It was used to obtain full-length human pancreatic
 CC islet GAD cDNA clones.
 SQ Sequence 32 BP; 6 A; 16 C; 6 G; 4 T;

 Query Match 100.0%; Score 18; DB 1; Length 32;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 DB 1 GACTCGAGTCGACATCGA 18

 RESULT 11
 ID Q36668 standard; cDNA: 35 BP.
 AC Q36668;
 DT 09-JUN-1993 (first entry)
 DE PPO adapter primer #1.
 KW Polyphenol oxidase; PPO; catalyst; browning; fruit; plastid; vacuole;
 KW transform; coffee; tea; black olives; grapevine; chloroplast; apple;
 KW transit peptide; recombinant plasmid; PCR; primer; amplify; broad bean;
 KW potato; polymerase chain reaction; ss.
 OS Synthetic.
 PN WO9302195-A.
 PD 04-FEB-1993.
 PF 16-JUL-1992; AU0356.
 PR 17-JUL-1991; AU-007248.
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PI Dry IB, Robinson SP;
 DR WPI: 93-058792/07.
 PT DNA encoding polyphenol oxidase polypeptide or fragment - useful
 PT for modifying the oxidase activity in fruit and vegetables to
 PT decrease or enhance browning
 PS Claim 15; Page 23; 44pp; English.
 CC The sequences given in Q36668-69 are adaptors which were used in the
 CC cloning of the polyphenol oxidase (PPO) enzyme genes from various
 CC plants. The PPO genes were isolated, and recombinant plasmids for
 CC transformation of plant cells were produced by PCR using the primers
 CC given in Q36678-78. PPO is thought to be the predominant catalyst in
 CC browning of fruit caused by injury or damage. PPO is localised in the
 CC plastids of plant cells whereas the phenolic substrates of the enzyme
 CC are stored in the plant cell vacuole. This compartmentation prevents
 CC the browning reaction from occurring unless the plant cells are damaged
 CC and the enzyme and the substrate are mixed. The PPO gene sequences
 CC could be used to construct synthetic genes which may be used to
 CC transform plants to decrease expression of the enzyme gene. In some
 CC instances, eg. coffee, tea, black olives etc., it is desirable to
 CC increase the level of PPO to produce desired levels of browning or
 CC changes in flavour compounds. The grapevine PPO gene codes for an
 CC additional 103 amino acids upstream of the N-terminus of the mature
 CC protein. This region has the properties of a chloroplast transit

CC peptide and is most likely responsible for targeting of the protein
 CC to be imported into the chloroplast and processed to produce mature
 CC PPO. Transformation of plants with this gene may therefore result
 CC in correct targeting and maturation of the grapevine PPO in other
 CC species and result in accumulation of active grapevine PPO enzyme in
 CC the plastids of these tissues.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

 Query Match 100.0%; Score 18; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 DB 1 GACTCGAGTCGACATCGA 18

 RESULT 12
 ID Q48416 standard; DNA: 18 BP.
 AC Q48416;
 DT 09-MAR-1994 (first entry)
 DE PCR primer for adenosine receptor partial cDNA (clone hval-3a).
 KW Adenosine receptor; expression system; cardiovascular disorders;
 KW renal disorders; neurological disorders; assay; ss.
 OS Synthetic.
 PN GB2264948-A.
 PD 15-SEP-1993.
 PF 05-MAR-1993; 004582.
 PR 13-MAR-1992; US-850701.
 PR 13-MAR-1992; US-850702.
 PR 13-MAR-1992; US-850707.
 PR 15-JAN-1993; US-005945.
 PA (MERL) MERCK & CO INC.
 DR WPI: 93-290442/37.
 PT Human adenosine receptor proteins - and expression systems,
 PT including cells for use in adenosine receptor binding assay
 PS Disclosure; Page 15; 65pp; English.
 CC Purified adenosine receptor proteins can be used in assays to screen
 CC for therapeutic human adenosine receptor agonists, antagonists or
 CC binding enhancers. These agonists, antagonists and binding
 CC enhancers may be used in the treatment of cardiovascular, renal and
 CC neurological disorders. Two copies of this primer were used to
 CC amplify partial cDNA of an adenosine receptor clone (hval-3a) for
 CC its subsequent insertion into an expression vector.
 SQ Sequence 18 BP; 5 A; 5 C; 5 G; 3 T;

 Query Match 100.0%; Score 18; DB 1; Length 18;
 Best Local Similarity 100.0%; Pred. No. 0.32;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 DB 1 GACTCGAGTCGACATCGA 18

 RESULT 13
 ID Q57547 standard; cDNA: 39 BP.
 AC Q57547;
 DT 20-MAY-1994 (first entry)
 DE Oligonucleotide ZC2487, an RNA -> cDNA primer.
 KW Gutamine:fructose-6-phosphate amidotransferase; insulin resistance;
 KW diabetes; screening; antagonists; GFAI; ss.
 OS Synthetic.
 PN WO9321330-A.
 PD 28-OCT-1993.
 PF 22-APR-1993; 003773.
 PR 22-APR-1992; US-872648.
 PR 05-NOV-1992; US-973330.
 PA (ZYMO) ZYMOGENETICS INC.

PI Mckernan PA, Mcknight GL, Ohara PJ, Sheppard PO;
 PI Smith RA, Vanness J;
 DR WPI: 93-351740/44.
 PT Human glutamine-fructose-6-phosphate amido transferase and
 PT corresp. DNA - for screening antagonists which inhibit insulin
 PT resistance, for use in diabetes patients
 PS Example 1: Page 61; 77pp; English.
 CC The sequence is a primer for the preparation of cDNA from RNA
 CC obtained from human pheochromocytoma cells. The cDNA was amplified by
 CC PCR to obtain the human glutamine:fructose-6-phosphate aminotransferase
 CC gene. The gene and its product are useful for screening for its
 CC antagonists which inhibit its exogenous activity. Such antagonists
 CC inhibit insulin resistance and are useful in the treatment of
 CC diabetes patients.
 CC See also Q57548-63 and Q51188.
 SQ Sequence 39 BP; 5 A; 5 C; 6 G; 21 T;

Query Match 100.0%; Score 18; DB 1; Length 39;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGACTCGACATCGA 18
 ||||||||||||||||
 DB 1 GACTCGACTCGACATCGA 18

RESULT 14
 ID 052503 standard; DNA; 35 BP.
 AC Q52503; (first entry)
 DE Helminth aminopeptidase H11-1 adaptor primer.
 KW Helminth; aminopeptidase; H11-3; H11-2; H11-1; integral membrane;
 KW alpha-amino acyl peptide hydrolase (microsome); multigene family;
 KW antigen; vaccine; parasite; human; virus; microbe; PCR; primer;
 KW polymerase chain reaction; amplification; Haemonchus contortus; ss.
 OS Synthetic.
 PN WO9323542-A.
 PD 25-NOV-1993.
 PF 07-MAY-1993; G00943.
 PR 08-MAY-1992; GB-009993.
 PA (AGRI-) AGRIC & FOOD RES COUNCIL.
 PI Graham M, Knox DP, Munn EA, Newton SE, Oliver JJ;
 PI Smith TS;
 DR WPI: 93-386574/48.
 PT New DNA encoding amino-peptidase from Haemonchus contortus and
 PT derived peptide(s) - useful in protective vaccines, for new
 PT vectors, transformed cells and oligonucleotide, for incorporation
 PT in virus or microbe
 PS Example: Page 32; 137pp; English.
 CC The sequences given in Q52503-11 are primers which were used to
 CC amplify cDNA clones of the helminth aminopeptidase genes H11-1, -2
 CC and -3. The amplified sequences encode H110D which is a protein
 CC doublet which shows homology to a family of integral membrane amino-
 CC peptidases. The differences between the amplified clones can be
 CC attributed to different mRNAs of the multigene family, and also to
 CC different variants of the H110D-encoding sequence being present at
 CC different stages of the life cycle, or in strains differing in
 CC geographical origin. Antigenic fragments of the aminopeptidases
 CC encoded by the H11 genes may be used in vaccines to stimulate immune
 CC response against helminth parasites in humans or other animals. These
 CC DNA sequences may be incorporated into a virus or microbe and used in
 CC a similar manner.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 18; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGACTCGACATCGA 18
 ..||||||||||||||
 DB 1 GACTCGACTCGACATCGA 18

DB 1 GACTCGACTCGACATCGA 18

RESULT 15
 ID 059015 standard; DNA; 35 BP.
 AC Q59015;
 AD 12-APR-1994 (first entry)
 DE RACE PCR adaptor-dr.
 KW Phosphonositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphonositides;
 KW ss; amplification; p110.
 OS Synthetic.
 PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphonositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1: Page 39; 146pp; English.
 CC An SGBAR-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with pSVneo. Total RNA was
 CC isolated from the SGBAR-1 and a cDNA library constructed. RACE PCR
 CC was performed by synthesizing first strand cDNA from random hexamers
 CC on SGBAR-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-
 CC dr as primers. Products were fractionated using an agarose gel. The
 CC DNA was isolated from the gel and subjected to PCR using oligo 2280 and
 CC adaptor as primers. A product of 350 bp was further sequenced.
 CC See also Q51155-6, Q59012-23 and Q57522-3.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 18; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGACTCGACATCGA 18
 ||||||||||||||||
 DB 1 GACTCGACTCGACATCGA 18

Search completed: September 13, 1999, 15:59:31
 Job time: 4798 sec

THIS PAGE BLANK (USP 12)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:56 ; Search time 1694.61 Seconds
(without alignments)
20.952 Million cell updates/sec

Title: US-09-325-095-18
Perfect score: 18
Sequence: 1 GACTCGAGTCGACATCGA 18

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database : 0

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	100.0	337	32	AA340759	AA340759 EST46034
2	15	83.3	360	24	D75133	D75133 CEIK0968F8
3	15	83.3	242	32	AA331366	AA331366 EST35191
4	15	83.3	535	39	AA897864	AA897864 NCM7A9T3
5	15	83.3	822	51	A1723540	A1723540 hcg15IE3.
6	14.8	82.2	390	36	C72810	C72810 C72810 R1ce
7	14.8	82.2	251	43	A1224703	A1224703 g958h10.x
8	14.4	80.0	323	20	T18855	T18855 a03001r Tes
9	14.4	80.0	360	35	C50284	C50284 C50284 Yuj1
10	14.4	80.0	494	39	AA846538	AA846538 a156f12.s
11	14.4	80.0	246	49	AV015355	AV015355 AV015355
12	13.8	76.7	312	21	D48281	D48281 R1C514415A
13	13.8	76.7	377	25	N81968	N81968 T9ESTZY6190
14	13.8	76.7	357	26	W46073	W46073 mc79f07.r1
15	13.8	76.7	412	26	W97957	W97957 mg05d09.r1
16	13.8	76.7	260	27	AA049420	AA049420 mj34a06.r
17	13.8	76.7	432	29	AA189627	AA189627 me90d09.r
18	13.8	76.7	332	30	AA198427	AA198427 mw46c01.r
19	13.8	76.7	335	30	AA237205	AA237205 mx18d03.r
20	13.8	76.7	425	30	AA238027	AA238027 mx16b10.r
21	13.8	76.7	402	30	AA245117	AA245117 mw28d02.r
22	13.8	76.7	447	30	AA245606	AA245606 mx01f01.r
23	13.8	76.7	451	30	AA253764	AA253764 mw03c07.r
24	13.8	76.7	426	30	AA254869	AA254869 me84907.r
25	13.8	76.7	434	30	AA272892	AA272892 va39h02.r
26	13.8	76.7	396	33	AA403552	AA403552 vdt8d01.r
27	13.8	76.7	332	33	AA412997	AA412997 vdb6f09.r
28	13.8	76.7	440	34	AA470223	AA470223 vdb0f05.r
29	13.8	76.7	420	34	AA495288	AA495288 fa02907.r
30	13.8	76.7	388	36	AA600575	AA600575 vm75f06.r
31	13.8	76.7	178	36	AA637510	AA637510 vt32d10.r
32	13.8	76.7	313	37	AA709591	AA709591 vt37f03.r
33	13.8	76.7	401	38	AA739275	AA739275 vv51f11.r
34	13.8	76.7	297	39	AA835087	AA835087 ak63a09.s
35	13.8	76.7	317	39	AA869732	AA869732 vq15b08.r
36	13.8	76.7	286	40	AA966805	AA966805 sg011a1.f
37	13.8	76.7	421	41	A1048842	A1048842 uc84a04.Y
38	13.8	76.7	807	41	A1068903	A1068903 mme0004C
39	13.8	76.7	366	42	A1130122	A1130122 SMOY33CAN
40	13.8	76.7	316	43	A1222065	A1222065 qh02c11.x
41	13.8	76.7	293	44	A1286399	A1286399 ul75h09.Y
42	13.8	76.7	482	46	A1399018	A1399018 NCM10A5T3
43	13.8	76.7	587	47	A1534610	A1534610 SD07520.5
44	13.8	76.7	426	48	A1545100	A1545100 fb70f03.Y
45	13.4	74.4	540	51	AU070229	AU070229 AU070229

ALIGNMENTS

RESULT 1
AA340759
LOCUS
DEFINITION EST46034 Fetal kidney II Homo sapiens cDNA 5' end similar to similar to protein phosphatase 2a, beta, 55 kDa, mRNA sequence.
ACCESSION AA340759
NID 91992998
VERSION AA340759.1 GI:1992998

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 337)
Adams M.D., Kerlavage A.R., Fleischmann R.D., Fulton R.A.,
Bult C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D.,
White O., Sutton G., Blake J.A., Brandon R.C., Man-Wai C.,
Clayton R.A., Cline T.R., Cotton M.D., Earle-Hughes J., Fine L.D.,
Fitzgerald L.M., Fitzhugh W.M., Fritchman J.L., Geohagen N.S.,
Glodek A., Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr.,
Kelley J.M., Kelley J.C., Liu L.-I., Marmaro S.M., Merrick J.M.,
Moreno-Palauques R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M.,
Phillips C.A., Ryder S.E., Scott J.L., Sauder D.M., Shirley R.,
Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li Y.,
Bednarek D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,
Dimke D., Feng D.-F., Ferrite A., Fischer C., Hastings G.A.,
He W.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K.,
Kozak D.L., Kunsch C., Hungjun J., Li H., Weissner P.S., Olsen H.,
Raymond L., Wei Y.F., Wing J., Xu C., Yu G.L., Ruben S.M.,
Dillon P.J., Fannon M.R., Rosen C.A., Haseltine W.A., Fields C.,
Fraser C.M. and Venter J.C.
*Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)
96026280
On Sep 12, 1996 this sequence version replaced gi:1406936.
Other ESTs: EST46035 THC125647
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699036
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/ldb/hgi/hgi.html>)
Seq primer: M13-21.
Location/Qualifiers
1..337
/organism="Homo sapiens"
/db_xref="ATCC (inhost):142407"
/db_xref="taxon:9606"
/clone_lib="Fetal Kidney II"
/dev_stage="fetus"
/note="Organ: Kidney; Vector: pBluescript KS-; Site_1:
XhoI; Site_2: EcoRI"
BASE COUNT 63 a 88 c 95 g 88 t 3 others
ORIGIN

Query Match 100.0%; Score 18; DB 32; Length 337;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTGAGTCGACATCGA 18
|||||
Db 4 GACTGAGTCGACATCGA 21
RESULT 2
L/5133/c 360 bp mRNA EST 14-DEC-1995
LOCUS CELK0968F Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk96f8 5', mRNA sequence.
ACCESSION D75133
NID 91120917
VERSION 91120917
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
AUTHORS Rhabditina; Rhabditoidae; Rhabditidae; Peleoderinae; Caenorhabditis.
1 (bases 1 to 360)
Kohara Y., Mitsuki H., Nishigaki A., Mochizuki T., Sugimoto A. and
Tabara H.
Toward an expression map of the C. elegans genome
Unpublished (1994)
On Sep 21, 1992 this sequence version replaced gi:276359.
TITLE
JOURNAL
COMMENT
Contact: Yuj1 Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykoha@dbj.nig.ac.j
Insert length: 1628 Std Error: 0.00
High quality sequence stop: 345.
Location/Qualifiers
1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone_lib="Yuj1 Kohara unpublished cDNA"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue=whole animal"
BASE COUNT 77 a 103 c 99 g 79 t 2 others
ORIGIN

Query Match 83.3%; Score 15; DB 24; Length 360;
Best Local Similarity 93.8%; Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CTCGAGTCGACATCGA 18
|||||
Db 249 CTCGAGTCGACATCGA 234
RESULT 3
L/51336 242 bp mRNA EST 21-APR-1997
LOCUS AA331366
DEFINITION EST35181 Embryo, 7 week, subtracted (total cDNA) II Homo sapiens
cDNA 5' end similar to similar to DNA polymerase III, mRNA
sequence.
ACCESSION AA331366
NID 91983609
VERSION AA331366.1 GI:1983609
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 242)

REFERENCE 1 (bases 1 to 242)
Adams M.D., Kerlavage A.R., Fleischmann R.D., Fulton R.A.,
Bult C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D.,
White O., Sutton G., Blake J.A., Brandon R.C., Man-Wai C.,
Clayton R.A., Cline T.R., Cotton M.D., Earle-Hughes J., Fine L.D.,
Fitzgerald L.M., Fitzhugh W.M., Fritchman J.L., Geohagen N.S.,
Glodek A., Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr.,
Kelley J.M., Kelley J.C., Liu L.-I., Marmaro S.M., Merrick J.M.,
Moreno-Palauques R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M.,
Phillips C.A., Ryder S.E., Scott J.L., Sauder D.M., Shirley R.,
Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li Y.,
Bednarek D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,
Dimke D., Feng D.-F., Ferrite A., Fischer C., Hastings G.A.,
He W.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K.,
Kozak D.L., Kunsch C., Hungjun J., Li H., Weissner P.S., Olsen H.,
Raymond L., Wei Y.F., Wing J., Xu C., Yu G.L., Ruben S.M.,
Dillon P.J., Fannon M.R., Rosen C.A., Haseltine W.A., Fields C.,
Fraser C.M. and Venter J.C.
*Initial assessment of human gene diversity and expression patterns

JOURNAL
MEDLINE
COMMENT
based upon 83 million nucleotides of cDNA sequence
Nature 377 (5547 Suppl), 3-174 (1995)
On Sep 12, 1996 this sequence version replaced gi:1404942.

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423

Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
source
Location/Qualifiers

1..242
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):132998"
/db_xref="taxon:9606"
/clone_lib="Embryo, 7 week, subtracted (total cDNA) II"
/dev_stage="embryo, 7 wks"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 45 a 71 c 75 g 45 t 6 others
ORIGIN

Query Match 83.3%; Score 15; DB 32; Length 242;
Best Local Similarity 88.2%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 2 ACTCGAGTCGACATCGA 18
|||||
DB 1 ACTCGAGTCGACATCGA 17

RESULT 4
AA897864 535 bp mRNA EST 12-APR-1998
LOCUS NC47A973 Mycelial Neurospora crassa cDNA clone NM7A9 5' end, mRNA
DEFINITION
ACCESSION AA897864
NID 93044297
VERSION AA897864.1 GI:3044297
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
Sordariales; Sordariaceae; Neurospora.
1 (bases 1 to 535)
Leonard, P.M., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,
Nelson, M.A., Mitchell, J., Arnijo, A.M., Bean, L., Bluyes, E.,
Cushing, T., Erett, A., Fleharty, M., Gorman, M., Judson, K.,
Miller, R., Ortega, J., Pavlova, I., Petea, J., Todisco, S.,
Trujillo, R., Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S.
and Natvig, D.O.
Expressed sequences from confidial, mycelial, and sexual stages of
Neurospora crassa
Fungal Genet. Biol. 21, 348-363 (1997)
97435349
On Jan 19, 1998 this sequence version replaced gi:2152245.

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Natvig, D.O./Nelson, M.A.
Department of Biology
University of New Mexico
Cascabel Hall, Albuquerque, NM 87131, USA
Tel: 505 277 3411
Fax: 505 277 0304
Email: ngevbiol@unm.edu
Deposited in GSDB at the National Center for Genome Resources with
accession GSDB:S:1146735

Seq primer: T3.
FEATURES
source
Location/Qualifiers

1..535
/organism="Neurospora crassa"
/strain="74A"
/db_xref="taxon:5141"
/map="4p16.3; 2; X"
/clone_lib="NM7A9"
/clone_lib="Mycelial"
/sex="Mating type A"
/tissue_type="Mycelium"
/dev_stage="Mycelium"
/note="2x sucrose for 24 hours. cDNA directionally cloned
into pBluescript SK(-) using the Uni-ZAP XR vector system
(Stratagene, La Jolla, CA)."
BASE COUNT 136 a 176 c 109 g 114 t
ORIGIN

Query Match 83.3%; Score 15; DB 39; Length 535;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACTCGAGTCGACATC 16
|||||
DB 40 ACTCGAGTCGACATC 54

RESULT 5
A1723540 822 bp mRNA EST 10-JUN-1999
LOCUS hcglis1E3.T7 Haemonchus contortus intestinal mRNA Haemonchus
DEFINITION contortus cDNA clone hcglis1E3.T7 T7 similar to 60S Ribosomal
protein L10, mRNA sequence.
ACCESSION A1723540
NID 95041869
VERSION A1723540.1 GI:5041869
KEYWORDS EST.
SOURCE Haemonchus contortus.
ORGANISM Haemonchus contortus.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Strongylida;
Trichostrongyloidea; Trichostrongylidae; Haemonchidae; Haemonchus.
1 (bases 1 to 822)
Jasmer, D.P., Myler, P.J. and Roth, J.P.
Haemonchus contortus intestinal EST Project
Unpublished (1999)
On Dec 20, 1995 this sequence version replaced gi:1133538.

Contact: Jasmer, D.P.
Department of Veterinary Microbiology and Pathology
Washington State University
Bustad 380, Pullman, WA 99164-7040, USA
Tel: 509-335-6040
Fax: 509-335-8528
Email: djasmer@vetmed.wsu.edu

FEATURES
source
Location/Qualifiers
1..822
/organism="Haemonchus contortus"
/db_xref="taxon:6289"
/clone_lib="hcglis1E3.T7"
/clone_lib="Haemonchus contortus intestinal mRNA"
/tissue_type="intestinal"
BASE COUNT 240 a 207 c 211 g 156 t 8 others
ORIGIN

Query Match 83.3%; Score 15; DB 51; Length 822;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CTCGAGTCGACATCG 17
|||||
DB 517 CTCGAGTCGACATCG 531

RESULT 6
LOCUS C72810 390 bp mRNA EST 22-SEP-1997
DEFINITION C72810 Rice panicle at flowering stage Oryza sativa cDNA clone
E2310.1A, mRNA sequence.
ACCESSION C72810
NID 92428347
VERSION C72810.1 GI:2428347
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE 1 (bases 1 to 390)
AUTHORS Sasaki, T. and Yamamoto, K.
TITLE Rice cDNA from panicle at flowering stage
JOURNAL Unpublished (1996)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasakigab@affrc.go.jp.
Location/Qualifiers
1.390
/organism="Oryza sativa"
/strain="Nippondare"
/db_xref="taxon:4530"
/clone="E2310.1A"
/clone_1lb="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering
stage"
BASE COUNT 89 a 105 c 86 g 108 t 2 others
ORIGIN

Query Match 82.2%; Score 14.8; DB 36; Length 390;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAGTCGAGTCGACATCGA 18
|| ||||| |||||
Db 283 GAGTCGAGTCGACATCGA 300

RESULT 7
LOCUS A1224703 251 bp mRNA EST 21-DEC-1998
DEFINITION A1224703 q998h10.x1 NCI-CGAP_Brl4 Homo sapiens cDNA clone IMAGE:1999171 3',
similar to gb:L22154 60S RIBOSOMAL PROTEIN L37A (HUMAN);, mRNA
sequence.
ACCESSION A1224703
NID 93807416
VERSION A1224703.1 GI:3807416
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 251)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1995 this sequence version replaced gi:811445.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
unknown library type
Insert Length: 205 Std Error: 0.00
Seq primer: -40UP from GIBCO.
Location/Qualifiers
1.251
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1999171"
/clone_1lb="NCI-CGAP_Brl4"
/sex="female"
/tissue_type="normal epithelium"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: breast; Vector: PAMPI; mRNA made from breast
epithelium, cDNA made by oligo-dT priming. Directionally
cloned. Size selected on agarose gel, average insert
size 500 bp. Primary library, non-amplified."
BASE COUNT 88 a 44 c 53 g 49 t 17 others
ORIGIN

Query Match 82.2%; Score 14.8; DB 43; Length 251;
Best Local Similarity 88.9%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAGTCGAGTCGACATCGA 18
|| ||| ||||| |||||
Db 153 GATTCGCGTCGACATCGA 136

RESULT 8
LOCUS T18855 323 bp mRNA EST 25-SEP-1996
DEFINITION T18855 a03001r Testis 1 Homo sapiens cDNA clone a03001 3' end, mRNA
sequence.
ACCESSION T18855
NID 9600897
VERSION T18855.1 GI:600897
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 323)
AUTHORS Pawlak, A., Tousseint, C., Levy, I., Bulle, F., Poyard, M., Barouki, R.
and Gueliaen, G.
TITLE Characterization of a large population of mRNAs from human testis
JOURNAL Genomics 26, 151-158 (1995)
MEDLINE 95301283
COMMENT On Sep 21, 1992 this sequence version replaced gi:276479.
Contact: Gueliaen G
Unité INSERM 99
INSERM
Unité INSERM 99, Hôpital Henri Mondor, 94010 Creteil, France
Tel: (33)149813530
Fax: (33)148980908
Email: gueliaen@infoblogen.fr
Seq primer: M13 reverse.
Location/Qualifiers
1.323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="a03001"
/clone_1lb="Testis 1"
/note="Vector: pGEM 5zf(+); Site 1: EcoRV; site 2: NotI;
mRNA was prepared from human testis of a 27 years old man.
cDNA was prepared using a 15mer oligo dT anchored by two
degenerated bases at its 3' end and containing a NotI site

at its 5' end. The cDNA was cloned between EcoRV and NotI sites of pGEM 5zf(+). The 3' end is at the NotI site. The EcoRV site is lost during the cloning procedure. cDNA corresponding to abundant species were eliminated from this library.

BASE COUNT 73 a 91 c 63 g 96 t

Query Match 80.0%; Score 14.4; DB 20; Length 323;
Best Local Similarity 93.8%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATC 16
|||||
Db 290 GACTCGAGTCGACATC 305

RESULT 9

LOCUS C50284 360 bp mRNA EST 11-SEP-1997
DEFINITION C50284 Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk479h2 5', mRNA sequence.

ACCESSION C50284
NID g2387537
VERSION C50284.1 GI:2387537
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Mochizuki,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394150.

FEATURES
source Contact: Yuj1 Kohara
Gene Library Lab
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
Location/Qualifiers
1..360
/organism="Caenorhabditis elegans"
/strain="Cbl489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk479h2"
/clone_lib="Yuj1 Kohara unpublished cDNA"
/note="dev stage-varied, sex=Hermaphrodite male,
tissue.type=whole animal"

BASE COUNT 77 a 104 c 97 g 82 t

Query Match 80.0%; Score 14.4; DB 35; Length 360;
Best Local Similarity 93.8%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CTCGAGTCGACATCGA 18
|||||
Db 261 CTCGAGTCGACATCGA 246

RESULT 10

LOCUS AA846538 494 bp mRNA EST 31-DEC-1998
DEFINITION a156f12.s1 Soares testis_NHT Homo sapiens cDNA clone IMAGE:1394351
3', mRNA sequence.
ACCESSION AA846538

NTD 92932678
VERSION AA846538.1 GI:2932678
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS I (bases 1 to 494)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL On Jan 19, 1998 this sequence version replaced gi:2151248.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILM at:
www.bio.lnl.gov/db/ftp/image/image.html

Insert Length: 592 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 483.

FEATURES

source Location/Qualifiers
1..494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21q"
/clone="IMAGE:1394351"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAAGTCGAGCGCCGCCCAATTTTGTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cots, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 130 a 127 c 85 g 152 t

Query Match 80.0%; Score 14.4; DB 39; Length 494;
Best Local Similarity 93.8%; Pred. No. 3.3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATC 16
|||||
Db 292 GACTCGAGTCGACATC 307

RESULT 11
LOCUS AV015355 246 bp mRNA EST 03-JUN-1999
DEFINITION AV015355 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA
clone 1110059C18, mRNA sequence.
ACCESSION AV015355
NID g4792347
VERSION AV015355.1 GI:4792347
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

SEQUENCE
AUTHORS

Euterbia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 246)
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Ito, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,
Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuna, H., Oda, H.,
Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,
Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y.,
Tomihaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
On Mar 20, 1998 this sequence version replaced gi:2980208.

TITLE
JOURNAL
COMMENT

Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-resetc.riken.go.jp
Thermotabilization and thermostabilization of thermostable enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome-rtc.riken.go.jp) for
further details.

FEATURES
source

1..246
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="6"
/clone="1110059C18"
/clone_lib="Mus musculus 18-day embryo C57BL/6J"
/sex="mixed"
/dev_stage="18-day embryo"
BASE COUNT 65 a 56 c 58 g 67 t
ORIGIN

Query Match 80.0%; Score 14.4; DB 49; Length 246;
Best Local Similarity 93.8%; Pred. No. 2.8e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CTCGAGTCGACATCGA 18
|||||
DB 29 CTCGAGTCGACATCGA 14

RESULT 12
D48281 312 bp mRNA EST 02-AUG-1995
LOCUS R1C514415A Rice green shoot Oryza sativa cDNA, mRNA sequence.
DEFINITION D48281
ACCESSION 5701990
NTD
VERSION D48281.1 GI:701990
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.

REFERENCE 1 (bases 1 to 312)
AUTHORS Sasaki, T., Miyao, A. and Yamamoto, K.
TITLE Rice cDNA from callus 1995
JOURNAL Unpublished (1995)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba

Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@agr.affrc.go.jp
High quality sequence stop: 297.
Location/Qualifiers
1..312
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone_lib="Rice green shoot"
/note="Green shoot (8 days old)"
BASE COUNT 53 a 99 c 88 g 71 t 1 others
ORIGIN

Query Match 76.7%; Score 13.8; DB 21; Length 312;
Best Local Similarity 88.2%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCG 17
|||||
DB 160 GACTCGAGTCGACGCG 176

RESULT 13
N81968/c 377 bp mRNA EST 10-SEP-1997
LOCUS T9ESTZy61g09.r1 TgrH Tachyzoite cDNA Toxoplasma gondii cDNA clone
DEFINITION t9zy61g09.r1 5', mRNA sequence.
ACCESSION N81968
NTD 91257721
VERSION N81968.1 GI:1257721
KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.

REFERENCE 1 (bases 1 to 377)
AUTHORS Hehl, A., Manger, I., Marra, M., Sidley, L.D., Ajlola, J.A.,
Aslett, M.A., Dietrich, N., Dubuque, T., Hillier, L., Kucaba, T.,
Man, K.L., Waterston, R.H. and Boothroyd, J.
TITLE WashU-Merck-Stanford-NIH Toxoplasma EST project
JOURNAL Unpublished (1996)
COMMENT On Sep 21, 1992 this sequence version replaced gi:276215.

Contact: Marra M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@wustl.edu
Clones can be obtained from Genome Systems Inc. (genome@MO.NET);
library can be obtained from Jim Ajlola (jw@molc.bio.cam.ac.uk)
Seq primer: T3
High quality sequence stop: 297.
Location/Qualifiers
1..377
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"
/clone="t9zy61g09.r1"
/clone_lib="TgrH Tachyzoite cDNA"
/note="Vector: lambda ZAP; site_1: EcoRI; site_2: XhoI;
Toxoplasma RH strain tachyzoites were grown in human
 foreskin fibroblast cultures in vitro. The library was
 synthesized by K.L. Man, Cambridge University. cDNAs were
 directionally cloned into the EcoRI to XhoI sites of the
 lambda zapII vector using the ZAP-cDNA synthesis kit.

(Stratagene). WARNING: the library contains a small
percentage of cDNAs derived from the human host cells.
BASE COUNT 102 a 78 c 111 g 84 t 2 others
ORIGIN

Query Match 76.7% Score 13.8; DB 25; Length 377;
Best Local Similarity 88.2%; Pred. No. 6.4e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 ACTCGAGTCGACATCGA 18
|||||
Db 340 ACTCGAGTCGACATCGA 324

RESULT 14
W46073 357 bp mRNA EST 23-MAY-1996
LOCUS m079507.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:354757 5' similar to PIR:S28237 S28237 NADH
dehydrogenase ;, mRNA sequence.
W46073
ACCESSION 91330807
NID W46073.1 GI:1330807
VERSION W46073.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 357)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
TITLE Unpublished (1996)
JOURNAL On Oct 18, 1995 this sequence version replaced gi:1023145.
COMMENT

CONTACT: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through ILNLT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:226557
Putative full length read
Seq primer: ETPRimer
High quality sequence stop: 343.
Location/Qualifiers
1. 357
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:354757"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue-type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGTGGAGCGCGCGGAAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo.
BASE COUNT 89 a 98 c 91 g 79 t
ORIGIN

Query Match 76.7% Score 13.8; DB 26; Length 357;
Best Local Similarity 88.2%; Pred. No. 6.3e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 ACTCGAGTCGACATCGA 18
|||||
Db 177 ACGCGTGTGACATCGA 193

RESULT 15
W97957 412 bp mRNA EST 16-JUL-1996
LOCUS m05009.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:422897 5' similar to PIR:S28237 S28237 NADH
dehydrogenase ;, mRNA sequence.
W97957
ACCESSION 91427866
NID W97957.1 GI:1427866
VERSION W97957.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 412)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
TITLE Unpublished (1996)
JOURNAL On May 9, 1995 this sequence version replaced gi:803042.
COMMENT

CONTACT: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through ILNLT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:257449
Seq primer: mob.REGA+ET
High quality sequence stop: 329.
Location/Qualifiers
1. 412
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="19"
/clone="IMAGE:422897"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue-type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGTGGAGCGCGCGGAAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and
M. Fatima Bonaldo.
BASE COUNT 111 a 108 c 100 g 93 t
ORIGIN

Query Match 76.7%; Score 13.8; DB 26; Length 412;
Best Local Similarity 88.2%; Pred. No. 6.5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 ACTCGAGTCGACATCGA 18
||| |||||
Db 167 ACCCGTGTGACATCGA 183

Search completed: September 13, 1999, 15:45:56
Job time: 4044 sec


```

/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PID:6304808"
/db_xref="PID:63294346"
/db_xref="GI:2294346"
/translation="MPRRSGEINGIHMPRIIVECILLPNCMTVTECLREANTLT
IKHELFKARKKYPHLOLDESSITVSTQAREEPEDETRRLDLEFPPLKVI
EYVGNKEIKILNREISFALGMPVCEPDVMDKDEVDIFRNITLNVCEAVDLNDSPH
SRAMVYPERNVESELPKRIHINLKDQGIIVYVYVSPNDKOKYTLKIHNDVPH
OYIAAIRKTRSMILSEBQLKCYEQGKILKYGCEDEFELEKPSQYKTRISC
IKLGNPNTLMKAKESLGSOLPMDCFTMSYSRISTAPYNGESTSLVYNSAL
RKLICATVYVNTIDIKYRTGIVHGCEPLCNVNTORPCSPRNENLNDIY
ITDLPRAATLCTISYKGRKAKHECHPLANGINLPTITLIVSGKALMLPEPH
GLEDLNPILGVTGSNPKETPCLELEDFESSVYKFPDVSIEBHANSVSRSEAGSY
SHAGLSNRLARDNLENDKEQLRAICTRDPLSEITEQKDLMSHRHCVTIPILP
KLLSVKNSRDVAMQYCLVADMPPIKPEQAMLELDCYPPPMVGFAYRCLKYLIT
DKLSOYLITOLVQLKVEQYLDNLVRLFKALTNORIGHEFMILKSEMKNVSO
REGLLSEYCRAGCAGYKILNVOVEMELINLTJDLKQEKDEOKYOMKRIYVOMR
RDPDMDALOGESPLNPAHQNLNREERINSKARPMLNENPDIHSELHPONE
ILFKGDDLRQMLTLQITIRIMENIMQNGDLRLMPLYCGLSIGCVGLIEVNSHT
IMQIOCKGLKGLQFNSTHLQWLKDNKGIBYDAIDLFTRSCAGYVATFIIIGIG
DHNSNIMYKDGOLFHDGFHDLKFKKFEYKREYVFTVLOFLIYISGAECEY
KTRFEREQEMCYKAVLAIROHANLEINFSMVLGSGMPELQSFODIAYIRRTLADK
TEQEALEYPMKMDAHGCGMTXMDMIFHTIKAHAN"
BASE COUNT      1147 a      626 c      722 g      1003 t
ORIGIN

```

Query Match

Best Local Similarity 100.0%; Score 20; DB 5; Length 3498;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1  TGCTGTAATTTCTAATGCTG 20
          11111111111111111111
Db      3325 TGCTGTAATTTCTAATGCTG 3306

```

```

RESULT  2
LOCUS   AC004707 123585 bp DNA PRI 23-JUN-1998
DEFINITION Homo sapiens chromosome 17, clone hRPC.117_B_12, complete sequence.
ACCESSION AC004707
NID       G3249127
VERSION  AC004707.1 GI:3249127
KEYWORDS HTG.
SOURCE   human.
ORGANISM Homo sapiens

```

```

REFERENCE 1 (bases 1 to 123585)
Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Birten, B., Fasman, K., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 17, clone hRPC.117_B_12
Unpublished
2 (bases 1 to 123585)

```

```

AUTHORS Birten, B., Fasman, K., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Baker, J., Baldwin, J., Barne, N., Beckerly, R., Benn, J., Boatin, C.,
Boutwell, C., Brown, A., Castle, A., Cerny, J., Cooke, P., Depayre, E.,
Devon, K., Dewar, K., Donelan, L., Etemadi, S., Ferreira, P.,
Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S.,
Gensheimer, S., Geraigery, K., Gilmartin, T., Grant, G., Hagos, B.,
Harris, K., Horton, L., Howland, J.C., Hul, L., Jacotot, L., Kann, L.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
Melchior, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Nachman, A., Nahf, R., Naylor, J., O'Connor, T., Pavlin, B.,
Peterson, K., Riley, R., Roberts, D., Rossello, R., Roy, A., Shyam, R.,
Stange-Thomann, N., Stillwell, J., Stojanovic, N., Stone, C.,
Strickland, C., Subramanian, A., Torruella-Miller, I., Vassiliev, H.,
Vo, A., Wagner, A., Wang, B., Wheeler, J., Wu, Y., Ye, W. J., Zhao, J. and
Zody, M.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 123585)

```

```

AUTHORS Birten, B., Fasman, K., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Baker, J., Baldwin, J., Barne, N., Beckerly, R., Benn, J., Boatin, C.,
Boutwell, C., Brown, A., Castle, A., Cerny, J., Cooke, P., Depayre, E.,
Devon, K., Dewar, K., Donelan, L., Etemadi, S., Ferreira, P.,
Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S.,
Gensheimer, S., Geraigery, K., Gilmartin, T., Grant, G., Hagos, B.,
Harris, K., Horton, L., Howland, J.C., Hul, L., Jacotot, L., Kann, L.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
Melchior, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Nachman, A., Nahf, R., Roberts, D., Rossello, R., Roy, A., Shyam, R.,
Peterson, K., Riley, R., Roberts, D., Rossello, R., Roy, A., Shyam, R.,
Stange-Thomann, N., Stillwell, J., Stojanovic, N., Stone, C.,
Strickland, C., Subramanian, A., Torruella-Miller, I., Vassiliev, H.,
Vo, A., Wagner, A., Wang, B., Wheeler, J., Wu, Y., Ye, W. J., Zhao, J. and
Zody, M.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jun 23, 1998 this sequence version replaced g1.3237292.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
source location/Ovalifiers
1..123585
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPC.117_B_12"
/clone_lib="Peter de Jong/ human PAC library"
/map="17"
/chromosome="17"
/complement(8..159)
/rpt_family="MIR"
216..479
/rpt_family="AluB"
1591..1853
/rpt_family="L1P3"
complement(1874..1995)
/rpt_family="FLAM_A"
2317..2615
/rpt_family="AluB"
complement(2756..2816)
/rpt_family="MER2"
2824..3127
/rpt_family="AluX"
3207..3264
/rpt_family="L1MC/D"
complement(3459..3624)
/rpt_family="MER2"
complement(4022..4322)
/rpt_family="AluB"
4738..4758
/rpt_family="AT-rich"
4959..5259
/rpt_family="AluSp"
5262..5562
/rpt_family="AluY"
5567..5595
/rpt_family="(CAA)n"
5774..5898
/rpt_family="FLAM_A"
complement(5928..5965)
/rpt_family="AT-rich"
5974..6280
/rpt_family="AluB"
6283..6406
/rpt_family="L1MB3"
6521..6822
/rpt_family="AluSp"
6825..6862
/rpt_family="L1MB3"
complement(6995..7285)
/rpt_family="AluB"
complement(7354..7657)

```


AAAPLPKPTAPVADVCETDLYDQGAALVEPRKKIKIIPQEGCVRGSTVI
SPRAVLRNQLEPIHELAQVAKIITHGGTGRISVPIAKVLLPGCPMSQHPAL
SSSALVYNEREFNLKRIHIAITKGAKNTEBQYKCAKDIEHYVDVADAKCVK
REHAQGLVYVGLINPPIHELAYEGITAPAYHEIETIGTPSGSKALIKSTVT
LMDLTSGKKNCKEIEIENDOKRMGTINTRTVSDLLNGMKAVDVLVDEPACHA
GTMALIAIVKPRKRVLGDPKPMFENLMOLKNNPREDICSTHYKISNRCT
OVRTATVSTLHIDGKRTNPKCKRAIEIDVNSYTKRKDIYLTCTRGWYKQGDIP
GGGHDRNASQGLTRKGVAVROKVENLIMKSHVAVLITREDRLVMTLQDDE
WIKYLVNKGNETALEMOQAEHEDIMAINSTSVVAVLITREDRLVMTLQDDE
TAGIELTPEQWEDLEPQFNDQPYSYVALDVICTMFGMDSSGIESPELPEEP
AVGKRVRAHNDSPGKFGKYNKAVIPTCKKPYVIRAKGQDILPIYGRVSPSRN
NIVPLNRNLPHSLITAGLOKKAAPLHKELNOLRPHSMILYSECTCYSKRTIWAFL
GYRGADHNDLHRGPPLSRDYLVYNNMCPYRPHGHOCEENHAGMRLIASALNCL
KRGGITALKATGAFDSDNSDVLSLAKFVRSVAVSFCTQTEMEFYKQDNDRE
ROFTHHLNLAVENTEDYKDSGAPSVYRVMNADCTEAVVANAARAKPGVYG
CAIEFKMKPSEENATEVEYAVMKCHKRVYHAAGPDEPRTLEAKRLKQNVHD
VAKIVENGESSVAIPILSTIGIAGADLDLSLRCFLTALRTDADVITICLDKKE
OHADAIRMBEYVTELDKPDIEIDELTAVHDSQCLKHIGYSTQYGRKLYSEGRKF
HOTADIAIRALPDPVQANQOICITGEPMESREKCPEDSPASAPRTIPELC
MAAMTIERICRHSNVITVCSPLPRIRIKNOQIQCKVLFNDVPIIPAR
VINDPEPVTPTDSDPTCSRLSLTITLNAEDIVSLTESEIDSLSLNEPAR
HVMISFKLRATYAIQALPQKLSMRREDRTPPPPPPPRPRAALSLANQIMELR
RHAITSVOAEVYHNSGFTPEALNERSGILRKPPVPLRQCTNLSRLANQIMELR
ITEGDAEGLDELRLTPTPTPEGDSQDSEOMERFEGNIGYVGGVIFSSDQGH
OOKSVYONTEILLERSLERIKHAPVLQKEMKCYQSPYANRSRQSKRYE
NNKATYTGRLDGLKATVPDYBACYKITTREPMISASVPREPSPEVAVACNNE
HENYPTVASIQLTDEYDALDWEGSSCLDTATPCPALRSKPHSLTEPLNSAV
PSAFONTLONVYLSAATKNCNVTOMRELVLDSAVNECFKYACNDYMEFEKRP
IRITTECVTSYARLAPKPEALALFAKTHQVLOEYPMOEFYMDKADYKVPGRKHT
EERPKVOVIOAAPLAKATAYIGIHELRLVRLTVALPINHTLDEMSAEDFQALIANF
SVYHPLETIDISFKSDSDSLATLALMLEDLGYDRLDEKAFEGEITVHLPATA
TFPKRGAMKSGMELTLENTVLYLVNVIASRVLEORLRDSCKAFITGDDITIGVSDK
IMADRCATWNNNEVKITIDAVIGIKAYIEGCGGILEOVVHTACRSDPKLRIKLGKP
LFDVDEQDHRRLALEDERTNANFRVGIQELRLKAESREYEOVQPVLLATLAFERSD
KAFKALGSPRLHYGPK"

7628..11362
/codon_start=1
/product="polypeptide 2"
/protein_id="A01362.1"
/db_xref="PID:g4240569"
/db_xref="GI:4240569"
/translation="MNSVFNPFGRGAVYAPPIAPRRRRAAPAPPSGLTQIOOLT
RAVRLVILNDRORPARTRPRKTKQPKPKONOKPOOQKKGKNOPOPKPK
PKRRORTAKPEADRTFVKNEDKMGAVAMEGVITPLVYKGTIDBPALAKLKT
KSSSTDMERAKLPTKMSDAFETTEHPEVFNHMGAVOFGSGRTTIFGVGPGDS
GRLDNSGRVAVIYAGANEVPTALSVTNKGAALIKTHEDIVEMSRATIMCI
LONVTPDCRPTCYNRNBDLTLMLETVNHSYVLLDALRCPTRRHVSSTPTDD
FTLAPYGLICHRCKTMEPCYSPRIKIEKWDADQDVLIOVSAOLGYRABATASAR
LRFMGGVPEIIOEGALIDEKVTSKPCLHSHKGVFVYVCRPGDSITTSKLGSD
OCTILPMRYGVKFGVERKTTLPNHTQIPCLITERTREKSGYTTMHPQOOSTML
MEESGEVYVPTSGNANVTECKCGDFKGTIVTARKIDGCTERQCLASDHAKW
FNSPLIRTHDTAQGLHIPPLOAOCTVPLALPGVNAVRSMSLTAHETLL
LTRHNGENQPTAEVYVSVTNFNSITIGFEYTWGNOKPVRVYAOESAPDHCHP
EARNIHVPYTOAPSGFEFNKNSGOLSTAPFGCVVNNRADCAVSGLPISVDI
PQAATRVSEPLSLIKCTVSTCTDVGGLVITVDRAGGACAVHSHSSTAVLD
PSVYEGKGETTLKSTRSLQADFEVSKGCTITTCACQCPTEHVMNRPQNSTPDS
SHKTSNMWITLALMGSSIAIAIVYVIALVFTPAQR"

BASE COUNT 3462 a 2906 c 2828 g 2628 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 11824;
Best Local Similarity 94.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 TCGTGAATCTAATGCT 19
|||||||

Db 11350 TCGTGAATCTAATGCT 11332

RESULT 4
FLAH38NHAH/c
LOCUS Influenza A virus gene for hemagglutinin precursor (hemagglutinin 1 region), partial cds, isolate: A/Equine/Rome/5/91.
DEFINITION
ACCESSION D30684
MID 9487670
VERSION 3.0684.1 GI:487670
KEYWORDS hemagglutinin precursor; HA1; hemagglutinin 1.
SOURCE Influenza A virus (strain:H3N8, isolate:A/Equine/Rome/5/91) cdna to genomic RNA.
ORGANISM Influenza A virus
VIRUSES: ssRNA negative-strand viruses; Orthomyxoviridae; Influenza virus A and B group; Influenza A viruses.

REFERENCE 1 (bases 1 to 1090)
Lindstrom, S.E., Endo, A., Pecoraro, M.R., Sugita, S., Damiant, A., Hiroto, Y., Kamata, M., Kumanomido, T., and Nerome, K.
Complete nucleotide sequence of the HA1 region of the hemagglutinin gene of A/Equine/Rome/5/91 (H3N8) influenza virus
Unpublished (1994)
Submitted (10-May-1994) to DDBJ by:
Stephen E. Lindstrom
Virology I
National Institute of Health
1-23-1 Toyama, Shinjuku-ku
Tokyo 162
Japan
Phone: 03-5285-1111 x2531
Fax: 03-5285-1155.

FEATURES
source location/Qualifiers
1..1090
/organism="Influenza A virus"
/isolate="A/Equine/Rome/5/91"
/strain="H3N8"
/db_xref="taxon:11320"
/chromosome="segment 4"
10..57
10..>1090
/codon_start=1
/product="hemagglutinin precursor"
/protein_id="BA38945.1"
/db_xref="PID:d1034920"
/db_xref="PID:g3779054"
/db_xref="GI:3779054"
/translation="MKTIILILTHWYSQNPSTSGNNTATLIGCHHAVANGTLYKTI
TDQLEVNNAELVQSIGIKICNNSTYVLDGRCTLIDALGPRCHDPRYENWDL
IERSASNCIPIDIPDIALRSIVASGTLFAGFTWGTGVONGSGCKRGKSGAD
SFFSLNLTSGNSYPLIANTMNNKFKDLTYGIIHPSNSNEQRLYIOESGRAT
VSTERQOCTVPIPNIGSRPLVVGOGGRISIVYIVKPGDILMNSNGNLVAPRGFKLR
TGKSSVMSQDLIDTCVSECTTPNGSIPNDKPFQNVKVTGYKCPKTIQWTLKANG
MRNPEKQIRLIFGALINGFWEG"

mat_peptide
BASE COUNT 391 a 217 c 232 g 250 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 1090;
Best Local Similarity 94.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 TCGTGAATCTAATGCT 19
|||||||

Db 420 TCGTGAATCTAATGCT 402

RESULT 5
HS105D16 148984 bp DNA PRI 30-NOV-1998
LOCUS HS105D16
DEFINITION Human DNA sequence from clone 105D16 on chromosome Xp11.3-11.4

Contains pseudogene similar to laminin-binding protein, CA repeat,
STS, complete sequence.

ACCESSION

AL031311

93947674

KEYWORDS

HTG, GI:3947674

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 148984)

AUTHORS

Pearce, A.

TITLE

Direct Submission

JOURNAL

Submitted (30-NOV-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

COMMENT

On Dec 2, 1998 this sequence version replaced gi:3647153.

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence is the entire insert of clone 105D16. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChX>
105D16 is from the library RPC16 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/VECTOR:RPC4>.

FEATURES

Source

1. 148984

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/clone="105D16"

/map="p11-3-11.4"

/clone_11b="RPC16"

1. 249

repeat_region /note="AluSp repeat: matches 59. .312 of consensus"

repeat_region /note="4 copies 12 mer 83% conserved"

repeat_region /note="1004. .1128

repeat_region /note="FLAM_C repeat: matches 1. .125 of consensus"

repeat_region /note="1617. .1914

repeat_region /note="AluSx repeat: matches 1. .299 of consensus"

repeat_region /note="2129. 2293

repeat_region /note="L1MC8 repeat: matches 1253. .1420 of consensus"

repeat_region /note="2535. .2839

repeat_region /note="AluSx repeat: matches 1. .305 of consensus"

repeat_region /note="2844. .3155

repeat_region /note="AluSx repeat: matches 1. .311 of consensus"

repeat_region /note="3161. 3454

repeat_region /note="HAL1 repeat: matches 547. .841 of consensus"

repeat_region /note="3207. .3472

repeat_region /note="L1MC8 repeat: matches 1669. .1937 of consensus"

repeat_region /note="3490. .4104

repeat_region /note="L1M4 repeat: matches 2370. .3002 of consensus"

repeat_region /note="4105. 4287

repeat_region /note="M1L1A2 repeat: matches 8. .201 of consensus"

repeat_region /note="4341. .4641

repeat_region /note="AluY repeat: matches 1. .298 of consensus"

repeat_region /note="5048. .5177

repeat_region /note="L1M4 repeat: matches 5666. .5809 of consensus"

repeat_region /note="5285. 5387

repeat_region /note="L1M4 repeat: matches 5434. .5530 of consensus"

repeat_region

5395. .5691 /note="AluSx repeat: matches 1. .297 of consensus"

repeat_region

5724. .6028 /note="AluSc repeat: matches 1. .308 of consensus"

repeat_region

6038. .6086 /note="AluSc repeat: matches 247. .295 of consensus"

repeat_region

6106. .6249 /note="L1M4 repeat: matches 5269. .5413 of consensus"

repeat_region

6363. .6629 /note="L1MC5 repeat: matches 7691. .7925 of consensus"

repeat_region

6876. .7193 /note="L1M4 repeat: matches 4477. .4785 of consensus"

repeat_region

7194. .7504 /note="AluSx repeat: matches 1. .311 of consensus"

repeat_region

7505. .7762 /note="L1M4 repeat: matches 4185. .4477 of consensus"

repeat_region

7763. .8097 /note="AluSg repeat: matches 1. .310 of consensus"

repeat_region

8098. .8205 /note="L1M4 repeat: matches 4081. .4185 of consensus"

repeat_region

8559. .8775 /note="HERV1 repeat: matches 5247. .5474 of consensus"

repeat_region

9055. .9080 /note="13 copies 2 mer ac 92% conserved"

repeat_region

9094. .9386 /note="AluSp repeat: matches 1. .294 of consensus"

repeat_region

9762. .10119 /note="THE1B repeat: matches 1. .362 of consensus"

repeat_region

10380. .10678 /note="AluSg repeat: matches 7. .306 of consensus"

repeat_region

10712. .11232 /note="L1M2 repeat: matches 5620. .6155 of consensus"

repeat_region

11218. .11621 /note="L1M2 repeat: matches 4955. .5351 of consensus"

repeat_region

11622. .11906 /note="AluSx repeat: matches 1. .286 of consensus"

repeat_region

11907. .12497 /note="L1M4 repeat: matches 4339. .4955 of consensus"

repeat_region

12494. .12543 /note="L1M4 repeat: matches 3102. .3154 of consensus"

repeat_region

12544. .12853 /note="AluSg repeat: matches 1. .306 of consensus"

repeat_region

12854. .13043 /note="L1M4 repeat: matches 2933. .3102 of consensus"

repeat_region

13070. .13224 /note="M1L1B repeat: matches 210. .390 of consensus"

repeat_region

14565. .14588 /note="12 copies 2 mer gt 96% conserved"

repeat_region

15584. .15895 /note="L1M8 repeat: matches 5507. .5860 of consensus"

repeat_region

15913. .16075 /note="L1M8 repeat: matches 6001. .6173 of consensus"

repeat_region

16092. .16394 /note="AluS8 repeat: matches 1. .302 of consensus"

repeat_region

16594. .16882 /note="AluSg repeat: matches 1. .296 of consensus"

repeat_region

17546. .17703 /note="M1R repeat: matches 61. .226 of consensus"

repeat_region

17736. .17993 /note="L1P8 repeat: matches 5889. .6148 of consensus"

repeat_region

18033. .18340 /note="AluSx repeat: matches 1. .310 of consensus"

repeat_region

18509. .18757 /note="AluSg repeat: matches 6. .271 of consensus"

repeat_region

19384. .19412 /note="M1R repeat: matches 236. .262 of consensus"

repeat_region

19413. .19624 /note="THE1B repeat: matches 171. .364 of consensus"

repeat_region

19625. .19928 /note="AluSx repeat: matches 1. .304 of consensus"

repeat_region

19929. .20091 /note="THE1B repeat: matches 3. .171 of consensus"

repeat_region

20092. .20280


```

repeat_region /note="MIR repeat: matches 30. .261 of consensus"
21765. .22188
/note="L2 repeat: matches 1687. .2104 of consensus"
repeat_region 22189. .22339
/note="BUR1 repeat: matches 1801. .1952 of consensus"
22575. .22880
/note="ALUSX repeat: matches 1. .306 of consensus"
25611. .25779
/note="MERSB repeat: matches 1. .171 of consensus"
27119. .27239
/note="L1MA5 repeat: matches 6180. .6300 of consensus"
27355. .27628
/note="L2 repeat: matches 2458. .2749 of consensus"
27657. .27963
/note="ALUdo repeat: matches 10. .312 of consensus"
28121. .28508
/note="L1MA10 repeat: matches 5979. .6319 of consensus"
28549. .29091
/note="L1ME1 repeat: matches 5624. .6165 of consensus"
29100. .29189
/note="45 copies 2 mer ta 71% conserved"
29240. .29628
/note="L1ME1 repeat: matches 5146. .5403 of consensus"
31238. .31401
/note="FRAM repeat: matches 1. .162 of consensus"
32068. .32179
/note="MIR repeat: matches 125. .235 of consensus"
32372. .32615
/note="MIR repeat: matches 1. .252 of consensus"
32853. .33009
/note="MER20 repeat: matches 1. .161 of consensus"
33187. .33283
/note="L2 repeat: matches 2584. .2700 of consensus"
33685. .33913
/note="MIR repeat: matches 31. .260 of consensus"
33918. .34091
/note="MIR repeat: matches 86. .260 of consensus"
34475. .34830
/note="MLR1C repeat: matches 3. .318 of consensus"
35067. .35214
/note="L2 repeat: matches 2585. .2723 of consensus"
35461. .35754
/note="ALUSG repeat: matches 1. .294 of consensus"
35830. .35959
/note="MIR repeat: matches 9. .142 of consensus"
36023. .36229
/note="L2 repeat: matches 2076. .2290 of consensus"
36348. .36709
/note="MLR1D repeat: matches 1. .398 of consensus"
38052. .38255
/note="L1MA9 repeat: matches 6101. .6308 of consensus"
38257. .39342
/note="L1G1R1 repeat: matches 2. .1164 of consensus"
39343. .41236
/note="L1PA2 repeat: matches 4227. .6146 of consensus"
41252. .42476
/note="L1G1R1 repeat: matches 1166. .2418 of consensus"
42479. .44326
/note="L1MA7 repeat: matches 4254. .6100 of consensus"
44314. .44903
/note="L1MAC repeat: matches 1467. .2414 of consensus"
44915. .45072
/note="L1PB3 repeat: matches 5919. .6090 of consensus"
45071. .45764
/note="L1M3E repeat: matches 1153. .1785 of consensus"
45765. .46081
/note="ALUtb8 repeat: matches 1. .318 of consensus"
46082. .46160
/note="L1M3E repeat: matches 1082. .1153 of consensus"
46161. .46587
/note="MSTA repeat: matches 1. .426 of consensus"
46588. .48395
/note="L1M3E repeat: matches -879. .1082 of consensus"

```

```

repeat_region 48598. .48800
/note="L2 repeat: matches 2541. .2750 of consensus"
48951. .49292
/note="L2 repeat: matches 1143. .1537 of consensus"
49553. .49661
/note="L2 repeat: matches 2089. .2195 of consensus"
49811. .49834
/note="L2 repeat: matches 2089. .2195 of consensus"
49837. .49905
/note="L2 repeat: matches 52. .117 of consensus"
49908. .49976
/note="MER58 repeat: matches 39. .108 of consensus"
50021. .50078
/note="29 copies 2 mer tt 78% conserved"
50079. .50113
/note="MER58 repeat: matches 2386. .2420 of consensus"
50501. .50743
/note="MIR repeat: matches 3. .262 of consensus"
54352. .54535
/note="L2 repeat: matches 2317. .2508 of consensus"
54858. .54934
/note="MLR1H repeat: matches 448. .525 of consensus"
55056. .55285
/note="MLR1H repeat: matches 4. .279 of consensus"
56812. .56820
/note="MER81 repeat: matches 1. .114 of consensus"
57445. .57502
/note="29 copies 2 mer tt 74% conserved"
57963. .58321
/note="L2 repeat: matches 2414. .2749 of consensus"
58856. .58943
/note="MIR repeat: matches 48. .139 of consensus"
59154. .59687
/note="L2 repeat: matches 149. .671 of consensus"
59714. .60013
/note="ALUSX repeat: matches 3. .302 of consensus"
60674. .61024
/note="L2 repeat: matches 1. .371 of consensus"
61221. .61728
/note="MLR1D repeat: matches 6. .505 of consensus"
62466. .62941
/note="HERV1 repeat: matches 3873. .4366 of consensus"
63839. .63878
/note="20 copies 2 mer gt 100% conserved"
65738. .70935

```

```

Query Match 84.0% Score 16.8; DB 10; Length 70935;
Best Local Similarity 90.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGCTGTAATTCATATGCTG 20
DB 30626 TGATGTAATTCATATGCTG 30607

```

```

RESULT 7
AC004468/C AC004468 127757 bp DNA PRI 10-NOV-1998
LOCUS Homo sapiens Xp21 pAC RPI1-37A12 containing exons 10 to 16 of the
DEFINITION Duchenne Muscular Dystrophy gene, complete sequence.
ACCESSION AC004468
VERSION 93097815
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 127757)
Muzny,D., Arenson,A.D., Adams,C., Bunac,C., Carvelli,K., Chang,J.,
Chacko,J., Chen,J., Ding,Y., Dugan,S., Durbin,J., Forcum,J.,
Ganesh,R., Garcia,C., Goodman,M., Gotrell,J.H., Haywood,M.,
Hernandez,J., Jackson,L., Jin,S., Kampal,R., Karpathy,S., Kovar,C.,

```

Lau, S., Leal, B., Lee, E., Li, Y., Licharge, O., Liu, M., Logan, O.,
 Lu, J., Ly, T., Marondei, I., Martinez, C., Meischer, S., Montgomery, K.,
 Oswal, G., Perez, L., Rashid, N.D., Renaldi, B., Rowland, K., Savage, L.,
 Scherer, S.E., Shen, H., Simon, M., Stovall, K., Timms, K.M., Todd, J.,
 Vo, O., Williamson, A., Worley, K.C., Yu, W., Kucherlapati, R.,
 Nelson, D. and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 127757)
 Worley, K.C.
 Direct Submission
 Submitted (21-MAR-1998) Molecular and Human Genetics, Baylor
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 127757)
 Worley, K.C.
 Direct Submission
 Submitted (01-MAY-1998) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 127757)
 Worley, K.C.
 Direct Submission
 Submitted (12-MAY-1998) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 5 (bases 1 to 127757)
 Gorrell, L.L.
 Direct Submission
 Submitted (31-OCT-1998) Human Genome Sequencing Center, Baylor
 College of Medicine, One Baylor Plaza, Houston, TX 77030, U.S.A.
 6 (bases 1 to 127757)
 Gorrell, L.L.
 Direct Submission
 Submitted (05-NOV-1998) Human Genome Sequencing Center, Baylor
 College of Medicine, One Baylor Plaza, Houston, TX 77030, U.S.A.
 7 (bases 1 to 127757)
 Gorrell, L.L.
 Direct Submission
 Submitted (06-NOV-1998) Human Genome Sequencing Center, Baylor
 College of Medicine, One Baylor Plaza, Houston, TX 77030, U.S.A.
 8 (bases 1 to 127757)
 Gorrell, L.L.
 Direct Submission
 Submitted (10-NOV-1998) Human Genome Sequencing Center, Baylor
 College of Medicine, One Baylor Plaza, Houston, TX 77030, U.S.A.
 On May 1, 1998 this sequence version replaced g1:3063478.
 Sequencing is completed to a minimum standard of double strand
 coverage with a minimum of 2 clones and 2 reads with no ambiguities
 or 2 chemistries with a minimum of 2 clones and 3 reads with no
 ambiguities. If the sequence quality does not meet this standard,
 it will be indicated in the annotation.
 The repeat regions shown were identified using RepeatMasker by
 Adrian Smit.
 Sequence similarities were identified using Powerblast by Jinghui
 Zhang.
 Exon/Intron boundaries of identified genes were chosen if there
 were canonical splice junctions that maintained sequence continuity
 across the splice junctions.
 FEATURES
 SOURCE
 1. 127757
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RPC11-37A12"
 /chromosome="X"
 /clone_lib="Roswell Park Cancer Center Human PAC Library"
 598. 663
 /rpt_family="CA)n"
 692. 757
 /rpt_family="MIR"
 repeat_region
 complement(882. 1384)

repeat_region
 complement(2098. 2156)
 /rpt_family="AT-rich"
 repeat_region
 complement(3257. 3506)
 /rpt_family="MIR"
 repeat_region
 6073. 6133
 /rpt_family="(CA)n"
 repeat_region
 complement(7015. 7069)
 /rpt_family="L2"
 repeat_region
 complement(10986. 11290)
 /rpt_family="Alusx"
 repeat_region
 11784. 11870
 /rpt_family="L2"
 repeat_region
 complement(11996. 12264)
 /rpt_family="L1MA5A"
 repeat_region
 complement(12308. 12533)
 /rpt_family="Alusg"
 repeat_region
 12534. 12947
 /rpt_family="L1PA6"
 repeat_region
 13190. 13267
 /rpt_family="L2"
 repeat_region
 13492. 13546
 /rpt_family="L2"
 repeat_region
 complement(13682. 13827)
 /rpt_family="MIR"
 repeat_region
 14135. 14573
 /rpt_family="L2"
 repeat_region
 14608. 14908
 /rpt_family="Aluv"
 repeat_region
 14937. 15275
 /rpt_family="L2"
 repeat_region
 15373. 15563
 /rpt_family="L2"
 repeat_region
 15701. 15752
 /rpt_family="AT-rich"
 repeat_region
 complement(17535. 17586)
 /rpt_family="AT-rich"
 repeat_region
 17677. 17858
 /rpt_family="MIR"
 repeat_region
 complement(20578. 20796)
 /rpt_family="MER58A"
 repeat_region
 22108. 22140
 /rpt_family="(CA)n"
 repeat_region
 complement(22136. 22182)
 /rpt_family="(CA)n"
 repeat_region
 23172. 27003
 /rpt_family="L1PA2"
 repeat_region
 complement(27357. 27463)
 /rpt_family="L1MA1"
 repeat_region
 27706. 28008
 /rpt_family="Alusg"
 repeat_region
 28011. 28031
 /rpt_family="Alusg"
 repeat_region
 28082. 28302
 /rpt_family="AT-rich"
 repeat_region
 complement(28304. 28359)
 /rpt_family="(CAA)n"
 repeat_region
 complement(29207. 29345)
 /rpt_family="MIR"
 repeat_region
 complement(29351. 29373)
 /rpt_family="AT-rich"
 repeat_region
 complement(29569. 29875)
 /rpt_family="Aluv"
 repeat_region
 complement(31436. 31550)
 /rpt_family="MIR"
 repeat_region
 31554. 32063
 /rpt_family="MER45B"
 gene
 join(33996. 34184,34835. 35016,64695. 64845,83272. 83391,
 105302. 105403,105511. 105618,113267. 113446)
 /gene="Homo sapiens dystrophin (DMD) mRNA, M18533."
 34507. 34655
 /rpt_family="L1R16C"
 misc_difference
 35335


```

/note="Insertion of C."
/clone="Insertion with respect to Y13187."
misc_difference 35340
/note="Deletion of G."
/clone="Deletion with respect to Y13187."
misc_difference 35358
/note="Deletion of A."
/clone="Deletion with respect to Y13187."
misc_difference 35386
/note="Deletion of A."
/clone="Deletion with respect to Y13187."
misc_difference 35413
/note="Deletion of G."
/clone="Deletion with respect to Y13187."
misc_difference 36366
/note="Deletion of T."
/clone="Deletion with respect to Y13187."
misc_difference 36893
/note="Deletion of CT."
/clone="Deletion with respect to Y13187."
misc_difference 36906
/note="Deletion of CT."
/clone="Deletion with respect to Y13187."
misc_difference 36980
/note="Deletion of C."
/clone="Deletion with respect to Y13187."
repeat_region /rpt_family="MAD1"
complement(37591..37670)
repeat_region /rpt_family="TAAn"
complement(37810..38338)
repeat_region /rpt_family="L2"
complement(38367..38448)
repeat_region /rpt_family="L2"
complement(38991..39104)
repeat_region /rpt_family="AT_rich"
complement(39105..39230)
repeat_region /rpt_family="FLAM_C"

Query Match      84.0% Score 16.8; DB 11; Length 127757;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGTCTAATCTAATGCTG 20
      ||||| ||||| ||||| |||||
Db 17931 TGCCTTAATTAATGCTG 17912

RESULT 8
AC004674 158886 bp DNA PRI 18-NOV-1998
LOCUS Homo sapiens Xp22 BAC GSHB-600G8 (Genome Systems Human BAC library)
DEFINITION complete sequence.
ACCESSION AC004674
VERSION 93273377
KEYWORDS AC004674.1 GI:3273377
SOURCE HTG.
ORGANISM human.
          Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 158886)
AUTHORS Morley,K.C.,
          Morley,K.C.,
          Carvelly,J.,
          Durbin,J.,
          Gorrell,J.H.,
          Kampal,R.,
          Liu,W.,
          Rashid,N.D.,
          Simon,M.,
          Morley,K.C.,
          Yu,W.,
          Chinault,C.,
          Nelson,D. and Gibbs,R.A.
          Submitted
          Title
          Journal
          Comment
          The repeat regions shown were identified using RepeatMasker by
          Adrian Smlt.
          Sequence similarities were identified using Powerblast by Jinghui
          Zhang.
          Exon/Intron boundaries of identified genes were chosen if there
          were canonical splice junctions that maintained sequence continuity
          across the splice junctions.
          Location/Qualifiers
            1..158886
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="GSHB-600G8"
              /chromosome="X"
              /clone_id="Genome Systems Human BAC library"
              /map="Xp22"
              complement(174..242)
              /rpt_family="MER5B"
              1904..1927
              /rpt_family="CAAAA)n"
              2946..3064
              /rpt_family="(GGA)n"
              6186..6364
              /rpt_family="MER58A"
              6380..6348
              /rpt_family="MER5B"
              6820..6982
              /rpt_family="FRAM"
              complement(8974..9232)
              /rpt_family="LIME1"
              9233..9529
              /rpt_family="AluXs"
              complement(9530..9807)
              /rpt_family="LIME7"
              complement(9817..9870)
              /rpt_family="LIME1"
              complement(9895..10264)
              /rpt_family="LTR16C"
              complement(10459..10770)
              /rpt_family="Aluub"
              12524..12641
              /rpt_family="MER86"
              12695..12749
              /rpt_family="LIME"
              14814..15117
              /rpt_family="AluXs"
              complement(15126..15240)
              /rpt_family="FRAM"

```

```

REFERENCE 2 (bases 1 to 158886)
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1998) Molecular and Human Genetics, Baylor
          College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 158886)
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-1998) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 158886)
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1998) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jun 30, 1998 this sequence version replaced gi:3264539.
          Sequencing is completed to a minimum standard of double strand
          coverage with a minimum of 2 clones and 2 reads with no ambiguities
          or 2 chemistries with a minimum of 2 clones and 3 reads with no
          ambiguities. If the sequence quality does not meet this standard,
          it will be indicated in the annotation.
          The repeat regions shown were identified using RepeatMasker by
          Adrian Smlt.
          Sequence similarities were identified using Powerblast by Jinghui
          Zhang.
          Exon/Intron boundaries of identified genes were chosen if there
          were canonical splice junctions that maintained sequence continuity
          across the splice junctions.
          Location/Qualifiers
            1..158886
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="GSHB-600G8"
              /chromosome="X"
              /clone_id="Genome Systems Human BAC library"
              /map="Xp22"
              complement(174..242)
              /rpt_family="MER5B"
              1904..1927
              /rpt_family="CAAAA)n"
              2946..3064
              /rpt_family="(GGA)n"
              6186..6364
              /rpt_family="MER58A"
              6380..6348
              /rpt_family="MER5B"
              6820..6982
              /rpt_family="FRAM"
              complement(8974..9232)
              /rpt_family="LIME1"
              9233..9529
              /rpt_family="AluXs"
              complement(9530..9807)
              /rpt_family="LIME7"
              complement(9817..9870)
              /rpt_family="LIME1"
              complement(9895..10264)
              /rpt_family="LTR16C"
              complement(10459..10770)
              /rpt_family="Aluub"
              12524..12641
              /rpt_family="MER86"
              12695..12749
              /rpt_family="LIME"
              14814..15117
              /rpt_family="AluXs"
              complement(15126..15240)
              /rpt_family="FRAM"

```

```

repeat_region      15590..15888      /rpt_family="AluSp"
repeat_region      15942..16947      /rpt_family="L1ME2"
STS               16236..16405
                  /db_xref="dbSTS:18691"
repeat_region      17043..17224      /rpt_family="L1ME2"
repeat_region      complement(17562..17860)
repeat_region      /rpt_family="AluUb"
repeat_region      complement(18976..19270)
repeat_region      /rpt_family="AluSx"
repeat_region      complement(19536..19589)
repeat_region      /rpt_family="(TA)n"
repeat_region      20280..20562
repeat_region      /rpt_family="AluIo"
repeat_region      22503..22802
repeat_region      /rpt_family="AluY"
repeat_region      23221..23448
repeat_region      /rpt_family="MIR"
repeat_region      complement(23594..23645)
repeat_region      /rpt_family="L1MA10"
repeat_region      complement(23637..23967)
repeat_region      /rpt_family="L1ME1"
repeat_region      23980..24049
repeat_region      /rpt_family="L2"
repeat_region      24383..24692
repeat_region      /rpt_family="AluUb"
repeat_region      25219..25342
repeat_region      /rpt_family="MIR"
repeat_region      25351..25780
repeat_region      /rpt_family="L1R165"
repeat_region      complement(25789..26097)
repeat_region      /rpt_family="AluIo"
repeat_region      26156..26215
repeat_region      /rpt_family="L2"
repeat_region      complement(26319..26587)
repeat_region      /rpt_family="L1MB4"
repeat_region      complement(27108..27516)
repeat_region      /rpt_family="L1"
repeat_region      27520..27660
repeat_region      /rpt_family="L1MB2"
repeat_region      27662..27953
repeat_region      /rpt_family="AluSx"
repeat_region      27963..28071
repeat_region      /rpt_family="L1MB2"
repeat_region      28072..28105
repeat_region      /rpt_family="MER4D"
repeat_region      28108..28162
repeat_region      /rpt_family="L1"
repeat_region      28163..28273
repeat_region      /rpt_family="AluUb"
repeat_region      28313..28367
repeat_region      /rpt_family="L1P5"
repeat_region      28368..28474
repeat_region      /rpt_family="(CATA)n"
repeat_region      28475..28774
repeat_region      /rpt_family="L1PB2"
repeat_region      28944..29091
repeat_region      /rpt_family="MER4D"
repeat_region      complement(29192..29314)
repeat_region      /rpt_family="MER34"
repeat_region      29338..29825
repeat_region      /rpt_family="L1MA9"
repeat_region      complement(30099..30327)
repeat_region      /rpt_family="MER20"
repeat_region      join(30331..30445,32077..32250,34228..34281,35716..35797)
repeat_region      36479..36517,36631..36696)
repeat_region      /gene="AA412296"
repeat_region      /note="AA412296 zui0h02.sl Soares_testis_NHT Homo sapiens
repeat_region      30715..31015

```

```

repeat_region      /rpt_family="AlusX"
                   31832..32162
                   /rpt_family="MER61"
                   complement(33303..33344)
repeat_region      /rpt_family="AT-rich"
                   33579..33688
                   /rpt_family="L2"
                   34905..35089
repeat_region      /rpt_family="AluY"
                   complement(36258..36422)
                   /rpt_family="FRAM"
                   37106..37402
repeat_region      /rpt_family="AlusX"
                   complement(39336..40010)
                   /rpt_family="L1P3"
                   40004..40559
repeat_region      /rpt_family="L1P47"
                   /rpt_family="L1P47"
                   40578..40612
repeat_region      /rpt_family="AT-rich"
                   40613..40634
repeat_region      /rpt_family="(GA)n"
                   complement(40674..40969)
                   /rpt_family="AluY"
                   51970..52439
misc_feature        /note="Region: AA402429 zt6be05.r1 Soares-testis-NHT Homo
                   sapiens CDNA clone 727520 5' similar to TR:G833928 G833928"
                   MJD1 PROTEIN"
                   52333..52863
misc_feature        /note="Region: Similar to AA746358 oa56f05.r1
                   NCI CGAP GCBI Homo sapiens CDNA clone IMAGE:1308993 5'"
                   53107..53141
repeat_region

Query Match      84.0%: Score 16.8; DB 11; Length 158886;
Best Local Similarity 90.0%: Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCTG 20
   ||||| ||||| ||||| ||
Db 3750 TGCTGTAATTCATGATG 3769

RESULT 9
AC005939 179485 bp DNA PRI 14-NOV-1998
LOCUS Homo sapiens chromosome 17, clone hRPK.467_K_17, complete sequence
DEFINITION AC005939
ACCESSION 93873180
NID 93873180
VERSION AC005939.1 GI:3873180
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 179485)
B1898.B.1 Linton.L., Nusbaum.C. and Lander.E.
Homo sapiens chromosome 17, clone hRPK.467_K_17
Unpublished
2 (bases 1 to 179485)
B1898.B.1 Linton.L., Nusbaum.C., Lander.E., Allen.N., Anderson.M.,
Baker.J., Baldwin.J., Barna.N., Beckerly.R., Benn.J., Boutwell.C.,
Brown.A., Castle.A., Cerny.J., Colangelo.M., Collins.S.,
Collamore.A., Cooke.P., Corliss.D., Depyre.E., Devon.K., Dewar.K.,
Donnell.L., Ferreira.P., Fitzhugh.W., Forrest.C., Funke.R.,
Gage.D., Gargana.S., Gerlanger.K., Grant.G., Hagoes.B., Heaford.A.,
Hersena.L., Horton.L., Howland.J.C., Jacotot.L., Jones.C., Kann.L.,
Karatsis.A., Lehoczy.J., Macdonald.P., Marquis.N., McEwan.P.,
McGurk.A., Mckenzie.K., Melidiri.J., Molla.M., Morris.W., Morrow.J.,
Mychalackey.J., Nahr.R., Naylor.J., Niloff.M., O'Connor.T.,
O'Donnell.P., Pavlin.B., Peterson.K., Riley.R., Roberts.D., Roy.A.,
Severy.P., Strange-Thomann.N., Stillwell.J., Stojanovic.N., Stone.C.,
Subramanian.A., Testaye.S., Tichovolsky.N., Tortorella-Miller.I.,
Vassiliev.H., Vo.A., Wagner.A., Wheeler.J., Wu.Y., Wyman.D.,
Ye.W.J., Zhao.J. and Zody.M.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

TITLE      Direct Submission
JOURNAL    Submitted (05-NOV-1998) Whitehead Institute/MIT Center for Genome
REFERENCE   Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS    3 (pages 1 to 179485)
            Baker,J., Baldwin,J., Barna,N., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donnell,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geradgery,K., Grant,G., Hagos,B., Heatford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczy,J., MacDonald,P., Margulis,N., McEwan,P.,
            McGurt,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nant,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torrella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.

FEATURES
SOURCE      Direct Submission
            Submitted (14-NOV-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Nov 14, 1998 this sequence version replaced g1:3868740.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
            Location/Qualifiers
            1. 179485
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="hbrp 467_K17"
               /clone_id="RPCT-11 human BAC library"
               /map="17"
               /chromosome="17"
               complement(917..1000)
               /rpt_family="L2"
               1100..1240
               /rpt_family="MER20"
               1404..1497
               /rpt_family="MER20"
               complement(1789..1988)
               /rpt_family="MIR"
               2070..2237
               /rpt_family="MIR"
               complement(2651..3041)
               /rpt_family="MSTID"
               complement(3128..3159)
               /rpt_family="(CA)n"
               3943..3964
               /rpt_family="AT-rich"
               4133..4175
               /rpt_family="(CAAA)n"
               complement(4821..5004)
               /rpt_family="MERSA"
               5017..6384
               /rpt_family="L1PA15"
               6413..6692
               /rpt_family="AluSc"
               6880..7033
               /rpt_family="L2"
               7034..7829
               /rpt_family="LTR1"
               7830..8129
               /rpt_family="L2"
               8183..8424
               /rpt_family="MIR"
               9028..9062
               /rpt_family="AT-rich"
               complement(9592..9891)
               /rpt_family="AluSc"
               10114..10209
               /rpt_family="MIR"
               complement(10551..10901)
               /rpt_family="L1ME3A"
               complement(11268..11491)
               /rpt_family="L1M4C"
               complement(13198..13412)
               /rpt_family="L1MA8"
               complement(13451..13697)
               /rpt_family="HERV16"
               13745..14036
               /rpt_family="AluYb8"
               complement(14738..15649)
               /rpt_family="Char1iel"
               15652..15948
               /rpt_family="AluY"
               complement(17202..17408)
               /rpt_family="AluSg/x"
               17653..17679
               /rpt_family="AT-rich"
               complement(17685..17979)
               /rpt_family="AluSc"
               18686..18802
               /rpt_family="MIR"
               19410..19544
               /rpt_family="L2"
               20448..21217
               /rpt_family="L2"
               complement(22467..22610)
               /rpt_family="MIR"
               complement(22839..22956)
               /rpt_family="L2"
               complement(23628..23675)
               /rpt_family="MLT2"
               complement(23724..24012)
               /rpt_family="MLT2"
               complement(24022..24387)
               /rpt_family="THE1B"
               complement(24388..24566)
               /rpt_family="MLT2"
               24676..25147
               /rpt_family="LTR33"
               complement(25261..25486)
               /rpt_family="AluSg/x"
               complement(25504..25936)
               /rpt_family="LTR8"
               complement(26042..26297)
               /rpt_family="LTR8"
               complement(26308..26453)
               /rpt_family="MIR"
               26484..26524
               /rpt_family="AT-rich"
               27458..27478
               /rpt_family="AT-rich"
               complement(27880..27921)
               /rpt_family="purine-rich"
               28488..28528
               /rpt_family="AT-rich"
               28660..29194
               /rpt_family="L1PA15"
               29195..29234
               /rpt_family="(TAAA)n"
               29589..29621
               /rpt_family="AT-rich"
               30085..30106
               /rpt_family="AT-rich"
               30553..30592
               /rpt_family="AT-rich"
               32459..32634
               /rpt_family="AluY"
               32642..32948
               /rpt_family="AluY"
               34130..34183
               /rpt_family="AT-rich"
               complement(34345..34382)
               /rpt_family="(CATA)n"

```

```

repeat_region complement(34755..35134)
                /rpt_family="MER37B"
repeat_region complement(35140..35725)
                /rpt_family="MER57-internal"
repeat_region complement(35993..36153)
                /rpt_family="HERV1"
repeat_region 36165..36199
                /rpt_family="(CATTT)n"
repeat_region complement(36200..36476)
                /rpt_family="AluJo"
repeat_region complement(36910..37139)
                /rpt_family="HERV16"
repeat_region 37162..37309
                /rpt_family="(TA)n"
repeat_region complement(37402..37687)
                /rpt_family="AluSp"
repeat_region 38963..38994
                /rpt_family="(GAAA)n"
repeat_region complement(39055..39216)
                /rpt_family="L2"
repeat_region 39848..39938
                /rpt_family="AT-rich"
repeat_region 39939..40222
                /rpt_family="AluSx"

```

```

Query Match      84.0% Score 16.8; DB 11; Length 179485;
Best Local Similarity 90.0% Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 TGCTGTAATTCATGCTG 20
    ||||| ||||| ||||| |||||
Db 162750 TGCTGAAATGCTATGCTG 162769

```

```

RESULT 10
LOCUS CELY73C8C 36277 bp DNA INV 05-MAR-1999
DEFINITION Caenorhabditis elegans cosmid Y73C8C.
ACCESSION AF101318
MID g3808329
VERSION AF101318.1 GI:3808329
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
Boulton, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kersey, J., Kirsten, D., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Ricken, L., Roopa, A.,
Saunders, D., Showkhen, R., Smaildon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sprat, J., and Wohlmann, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE
AUTHORS Bemis, G., Courtney, L., and Wohlmann, P.
TITLE The sequence of C. elegans cosmid Y73C8C
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 36277)
TITLE Waterston, R.
JOURNAL Direct Submission
AUTHORS Submitted (26-Oct-1998) Department of Genetics, Washington
REFERENCE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
AUTHORS 4 (bases 1 to 36277)
Waterston, R.

```

TITLE JOURNAL COMMENT

Direct Submission
Submitted (05-MAR-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rwenematode.wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

This is a portion of YAC Y73C8, sequenced to span the gap between
C50H11 and F02C9. The 5' clone is C50H11, 1561 bp overlap; 3' clone
is F02C9, 4400 bp overlap. Actual start of YAC Y73C8 is at base
position 37818 of CELC50H11; actual end is at 41552 of CELT28A11

NOTES:

Coding sequences below are predicted from computer analysis, using
the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES SOURCE

```

1..36277
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/clone="Y73C8C"
/chromosome="V"
complement(2834..4971)
/gene="Y73C8C.4"
complement(join(2834..3145,3402..3668,3755..4102,
4511..4684,4735..4971))
/gene="Y73C8C.4"
/codon_start=1
/evidence=not-experimental
/protein_id="AAC69342.1"
/db_xref="PID:g3808331"
/translation="MRLSICWVPLETGFNYIESKSTKLNSLNVCTOSEKAEINAK
TPYNNMLNENSIQVTEINKINNTGKIPTEGQPCNTEAOKNKEVRSKCE
KIPFYPMDCVSKYIALYRGYPCSGFNPLTRKDPDSRRKYSEBVRKIALR
TKDQKALEIDSTIERVDLSTKTHSHSLFHLVSLRCQPVADKMEETILKSEK
KSSLELTHTTEKTEAWTFEKKRWKSPISVDELDGKPLSPVQASHPEPARLRAPL
VCSEVVRVYIMKRPTRSRFLAALYMRGCGEAGATAEAGADLMDGADLKFGNMH
FOFPADFCMKSSYFSEDRFTNTEVLCNVYEDVFNCLNEVRTREDAKTPEFIDE
KFHPDONTTELKKECKOYVNGMCGDGLITLNDIMNDYQKNTNTEGSLI"
complement(5950..8732)
/gene="Y73C8C.5"
complement(join(5950..6008,6067..6515,6569..6944,
7624..7750,7803..8178,8242..8732))
/gene="Y73C8C.5"
/note="contains similarity to C. elegans olfactory
receptor ODR-10-related proteins"
/codon_start=1
/evidence=not-experimental
/protein_id="AAC69341.1"
/db_xref="PID:g3808330"
/db_xref="GI:3808330"

```

```
/translation="MSALTFLELSHVETIVEFEASTENMLILITFAVYKNGKRGTK
RMVYFSTLCGVSALEETAKPFSHNTNGSLITSTSELPETPFLMSMSATITIV
VATIAVOFVNTLCTINMATKFDGKNTVWIIIPLPAGFAVITISMFELCAAEYSO
SVYATISKNGLEVSKLPRYSMTASDQSLRNTYITLIGISISFHFITIVCC
YKMHISMKIELAKSVKQOLHROFNFALIVQIIGLPIITPIEFALIPILK
IDVMOGNTLALNIKOSLANSFGSINTIMIFKCFREIVRGGEGGNSYSATIPY
KRMIVFAAIGLMSGFEALSKPETHNNGTLFSAANPOLINISATFVNSCEY
LITTSFISVFIYRYLCLFDSKTRPFEGFGGIMLPLPGICYSTAITYLLAPND
YTDERVDSVAENGLNSVSELAFLSPYNSDSVIVHSIIFLLIASSLIHFHSIIL
FCSLKHMFNMKLEKFSVQNLQROFNLVQALVPMFLITPAVPLIAPFLAP
IVGIEVDWQSGNLSLVGFYPPCDICISVALVSEYRNLIKKNKISMCCNNNAOVSVINS
L"
gene
complement(9050.10372)
/gene="Y7308C.6"
complement(join(9050.9126,9174.9622,9885.10372))
/gene="Y7308C.6"
/notes="contains similarity to C. elegans olfactory
receptor ODR-10-related proteins"
cds
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69346.1"
/db_xref="PID:93808335"
/db_xref="GI:3808335"
/translation="MILYDLAGVFTKIOFCATFLINSFLILITCYIKVGAAYKRW
IYFAIGMTISGLEVTAPRYTHNKASLFTSVTNTPREVITCLMISMGCVITIV
SFIVQVYRYLCLFKAOKIKKFGVSKISFLIIPMGVGYTVAANLCLADYAD
YRGLVSEKSMELISNLPREPMLPYADGSRMSLVPLICIFLIGIHIILIIYCC
RMHSMKLEKLFSTONKLOFOFFHLLVOTGLTPIELVIPASDILVPLAPMFGF
DIDOTGNLFAFVGFPFPPDISIAFMIIYKEVYKIKKHVLCGLGTDNQVAVEAFON
VQS"
gene
complement(10614.12486)
/gene="Y7308C.7"
complement(join(10614.10877,10978.11266,11315.12486))
/gene="Y7308C.7"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69347.1"
/db_xref="PID:93808336"
/db_xref="GI:3808336"
/translation="MCHLELVKMOEDMDVSLKEVLAHPKLTFTYOEVARYSK
DKRIEYVIEVTAQFSAATPIITPGGSCVAVLPYVNLATNMTAKFERMD
SOFLIEEFEPDQYDTLRTGVMFIDIKVDDELNEKMETLIGLITCEPPTROK
GKGFSTSENFANLRLKLDIFPDICQSSITFEKIKAKSKLSTIGAGGLIHOLO
INCLARKIPKISERTHSOLACSDITIGCCICSDAUPRSMNEPVAVKISSDVSH
AELBEMETKLELSRSNEKTOEPENDPKLSKSKNRKSSQITAKPEKDSNA
PKCFRASEYKIKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK
LQAOIYVAKNHGNSRIPGEKIEPIRSAGETENVQILHEMLETOKLIEKNPMSK
ITMTSKLIEVSQNRNKLAKRELIIYERQIRGYTKAYEETFEKAPLILCAOLIKP
EPNDELNDICLDEKKEHEHTLKCDCRKRHTGCAQKMPVYKRICPCDAGLDD
DNFRPALS"
gene
complement(13005.16676)
/gene="Y7308C.8"
complement(join(13005.13704,13754.13837,13890.13976,
14030.15210,15770.16051,16111.16347,16416.16676))
/gene="Y7308C.8"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69348.1"
/db_xref="PID:93808337"
/db_xref="GI:3808337"
/translation="MFLNPTIEFTISPSLKLKLYCNKIPBELPLENDCTOKNPKNI
QNLIDSSNGKLVVGSABEITFESLNLFRFPGSHKFEFNAENDPAVQSGMKRVPYI
YKHYFVYIQAICVSPLRRETEVALATIASYKIRLILPLIHEVNPFOOTSVYPI
QOGRMNFNRTIGESLOLKNEYTEFEEMQMKISKFEMSSVHDSLOVLRFAK
KIPRDNELVYRMPANSKIVLEVERKIEHONMVEYRPEKNEYKYPVAFIRDPK
NKLIMHLLVYVMOHMDVSSFEKEVLEMKIATNGOEVAARVYKSPDEPNIEVIT
PIYAKYATPIPIITBGLCILASDFNLLVNLTKKIFOKIDBQFOLIREFFN
FIDFSGAKGVNFDLNDVEYINKMAEATKTLITISEEKAVNNIKTETEDLT
NELYKLDLIESEICQSDYFDMHTRAKMLKLPNVEFGAIVHLIDICRISLT
IPAFIHOLACSRIFGYRCPVADANTKADVAPPEELVEVEISPAVNEPETS
ESKRVKSKKOKRKPENIAPKEKQGNACPKGRASEYVNLKADDLRLEKIEKHL
RKSILSQOSESEKSKTAGKNEIEHAKOVELTRKELLDOKSLINKEKPIRLQK
EOLIDHONALVKKSIVIERDEBIOLLAQIOAOONALIAQOKISIRNLELNKLMOK
```



```

gene
9280.11907
/gene="BB0738"
9280.11907
/gene="BB0738"
/ncore="Similar to GB:M16318 SP:P11931 PID:143797 percent
identity: 47.36; Identified by sequence similarity;

Query Match
Best Local Similarity 94.4%; Pred. NO.3e+02; Length 11985;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CTGTAAATCTTAATGCTG 20
|||||
Db 10789 CTGTAAATCTTAATGCTG 10772

RESULT 12
HS424L16
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 107484)
Mashreqhi-Mohammadi M.
Direct Submission
Submitted (03-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 7, 1999 this sequence version replaced gi:4584767.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.

FEATURES
Source
1..107484
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="424L16"

BASE COUNT 33899 a 18771 c 20171 g 34643 t

ORIGIN

Query Match
Best Local Similarity 94.4%; Pred. NO.2.2e+02; Length 107484;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CTGTAAATCTTAATGCTG 20
|||||
Db 11101 CTGTAAATCTTAATGCTG 11118

RESULT 13
AC004098
LOCUS
DEFINITION
ACCESSION
NID
VERSION
AC004098.1 GI:3097872;


```

KEYWORD	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 166978)
JOURNAL	Barren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E. Homo sapiens chromosome 17, clone HCTR305D20 unpublished
AUTHORS	2 (bases 1 to 166978)
REFERENCE	Barren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckery,R., Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J., DePatre,E., Devon,K., Dewar,K., Durette,B., Etemadli,S., Ferrario,P., Forrest,C., Gage,D., Gardyna,S., Gensheimer,S., Geraldery,K., Gilmartin,T., Gray,D., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hul,L., Jacotot,L., Linton,L., McKenzle,J., Margulis,N., McEwan,P., McCurt,A., Meldrim,J., Mollin,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Pavlin,B., Peterson,K., Ranganathan,S., Riley,R., Roberts,D., Rollins,G., Rossello,R., Roy,A., Shyam,R., Soohoo,S., Stillwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Ye,W.J., Zemtseva,I., Zhao,J. and Zody,M.
TITLE	Direct Submision
JOURNAL	Submitted (01-FEB-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	3 (bases 1 to 166978)
AUTHORS	Barren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Boatin,C., Boutwell,C., Brown,A., Byrne,S., Cantu,C., Castle,A., Cerny,J., Cooke,P., DePatre,E., Devon,K., Dewar,K., Donelan,L., Durette,B., Etemadli,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S., Geraldery,K., Gilmartin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hul,L., Jacotot,L., Kann,L., MacDonald,P., Margulis,N., McEwan,P., McGurt,A., McKernan,K., Meldrim,J., Mollin,M., Morris,W., Morrow,J., Mychalecky,J., Nachman,A., Nahf,R., Naylor,J., O'Connor,T., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stillwell,J., Stone,C., Strickland,C., Subramanian,A., Torrella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
TITLE	Direct Submision
JOURNAL	Submitted (01-MAY-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On May 1, 1998 this sequence version replaced gi:3095021. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html .
FEATURES	Location/Qualifiers
SOURCE	1.166978
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="HCTR305D20"
	/clone_lib="Research Genetics/Cal Tech CTRB-HSP-C (plates 195-384)"
	/map="17"
	/chromosome="17"
repeat_region	complement(727..1166)
repeat_region	/rpt_family="MTJC"
repeat_region	complement(1788..1893)
repeat_region	/rpt_family="LIMC4"
repeat_region	1934..1978
repeat_region	/rpt_family="CAIN"
repeat_region	complement(2055..2357)
repeat_region	/rpt_family="AUSC"
repeat_region	complement(3360..3660)
repeat_region	/rpt_family="AUSX"
repeat_region	complement(3854..4144)
repeat_region	/rpt_family="AUSG"
repeat_region	complement(4191..4414)
repeat_region	/rpt_family="AUSB"
repeat_region	complement(4435..4621)

```

repeat_region /rpt_family="L1ME"
complement(4741..5042)
repeat_region /rpt_family="AluSp"
5273..5569
repeat_region /rpt_family="AluSx"
6359..6391
repeat_region /rpt_family="AT_r1ch"
complement(6650..6707)
repeat_region /rpt_family="(CA)n"
complement(7730..8031)
repeat_region /rpt_family="AluY"
8784..9067
repeat_region /rpt_family="AluUb"
complement(9222..9255)
repeat_region /rpt_family="AT_r1ch"
complement(9356..9554)
repeat_region /rpt_family="AluSx"
complement(9561..9685)
repeat_region /rpt_family="AluSg/X"
complement(9686..9764)
repeat_region /rpt_family="MIR"
complement(9764..9903)
repeat_region /rpt_family="L2"
10276..10360
repeat_region /rpt_family="L2"
10410..10582
repeat_region /rpt_family="MIR"
complement(11785..11874)
repeat_region /rpt_family="MER33"
complement(11881..1236)
repeat_region /rpt_family="L1MD3"
12578..12621
repeat_region /rpt_family="AT_r1ch"
complement(13069..13199)
repeat_region /rpt_family="L1MA7"
13238..13280
repeat_region /rpt_family="AT_r1ch"
14239..14330
repeat_region /rpt_family="AT_r1ch"
complement(15960..16041)
repeat_region /rpt_family="AT_r1ch"
complement(16384..16446)
repeat_region /rpt_family="AT_r1ch"
17101..17400
repeat_region /rpt_family="AluY"
complement(19091..19129)
repeat_region /rpt_family="AT_r1ch"
19192..19288
repeat_region /rpt_family="AT_r1ch"
19651..19766
repeat_region /rpt_family="AluUb"
complement(19767..19814)
repeat_region /rpt_family="AT_r1ch"
complement(20491..20712)
repeat_region /rpt_family="MIR"
complement(20819..20890)
repeat_region /rpt_family="MER20"
21092..21127
repeat_region /rpt_family="AT_r1ch"
complement(21616..21748)
repeat_region /rpt_family="MER20"
22002..22031
repeat_region /rpt_family="AT_r1ch"
22541..23399
repeat_region /rpt_family="L2"
23656..23730
repeat_region /rpt_family="L2"
complement(24469..24780)
repeat_region /rpt_family="AluDo"
24847..24877
repeat_region /rpt_family="AT_r1ch"
complement(27965..28156)
repeat_region /rpt_family="L1ME"

```

```

repeat_region 28669..28720
repeat_region /rpt_family="L2"
repeat_region complement(28816..28870)
repeat_region /rpt_family="POLY_A"
complement(28934..29213)
repeat_region /rpt_family="AluSx"
29355..29358
repeat_region /rpt_family="MIR"
complement(29583..29628)
repeat_region /rpt_family="AT_r1ch"
31312..31370
repeat_region /rpt_family="MER47A"
complement(31362..31438)
repeat_region /rpt_family="MER47B"
complement(32276..32458)
repeat_region /rpt_family="AluSg/X"
complement(32461..32760)
repeat_region /rpt_family="AluY"
complement(32808..33024)
repeat_region /rpt_family="AluUb"
33969..34139
repeat_region /rpt_family="MER5A"
34934..35227
repeat_region /rpt_family="AluUb"
35433..35718
repeat_region /rpt_family="AluSx"
35719..35768
repeat_region /rpt_family="AT_r1ch"
complement(37095..37519)
repeat_region /rpt_family="L1MA4A"
complement(37903..38084)
repeat_region /rpt_family="MIR"
38440..39130
repeat_region /rpt_family="L2"
39405..39567
repeat_region /rpt_family="AluDo"
42638..42663
repeat_region /rpt_family="AT_r1ch"
complement(44243..44397)
repeat_region /rpt_family="L1PB2"
complement(44447..44658)
repeat_region /rpt_family="L1PB2"
44669..44749
repeat_region /rpt_family="(TA)n"
complement(44757..44838)
repeat_region /rpt_family="AluUb"
45894..46277
repeat_region /rpt_family="MLT2CB"
complement(46281..46325)

```

Query Match 82.0%; Score 16.4; DB 11; Length 166978;
 Best Local Similarity 94.4%; Pred. No. 2.1e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CTGTAATCTAATGCG 20
 Db 164154 CTGTAATTCAAATGCG 164171

RESULT 14
 AC005968/c AC005968 145603 bp DNA PRI 22-NOV-1998
 LOCUS Homo sapiens chromosome 18, clone hRPK.24_A.23, complete sequence.
 DEFINITION AC005968
 ACCESSION AC005968
 NID 93907448
 VERSION AC005968.1 GI:3907448
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 145603)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.


```

repeat_region complement(33041..33537)
repeat_region /rpt_family="MIR2FB"
repeat_region 33731..33855
repeat_region /rpt_family="MIR5B"
repeat_region 33975..34069
repeat_region /rpt_family="LIME"
repeat_region complement(34102..34137)
repeat_region /rpt_family="(CA)n"
repeat_region 34138..34469
repeat_region /rpt_family="LIME"
repeat_region 34780..34810
repeat_region /rpt_family="(CAA)n"
repeat_region 34837..34871
repeat_region /rpt_family="POLY_A"
repeat_region 34889..35221
repeat_region /rpt_family="LIME"
repeat_region 35290..35679
repeat_region /rpt_family="LIME2"
repeat_region 35680..35956
repeat_region /rpt_family="AluSg1"
repeat_region 35957..35979
repeat_region /rpt_family="(CAA)n"
repeat_region 35981..36581
repeat_region /rpt_family="LIME2"
repeat_region 36698..37964
repeat_region /rpt_family="LIME2"
repeat_region complement(37965..38274)
repeat_region /rpt_family="AluY"
repeat_region 38275..38836
repeat_region /rpt_family="LIME2"
repeat_region complement(38858..39349)
repeat_region /rpt_family="L2"
repeat_region 39533..39691
repeat_region /rpt_family="MIR"
repeat_region complement(40792..41086)
repeat_region /rpt_family="MER33"
repeat_region complement(41227..41324)
repeat_region /rpt_family="L2"
repeat_region 42393..42450
repeat_region /rpt_family="AT-rich"
repeat_region complement(42660..42772)
repeat_region /rpt_family="LTR40a"

Query Match      82.0%; Score 16.4; DB 11; Length 145603;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      3 CTGTAATCTAATGCTG 20
        |||||||||||||||
Db 103680 CTTTAATCTAATGCTG 103663

RESULT 15
G09637/c      387 bp      DNA      STS      14-ADG-1995
LOCATION      human STS CHLC.GATA47F01.P15242 clone GATA47F01.
ACCESSION      G09637
VERSION      9941486
KEYWORDS      G09637.1 GI:941486
SOURCE      STS sequence; primer: sequence tagged site.
              human vector-pUC1 host-E.coli dutung+ (DH10B) Marker selected
              genomic DNA prepared from XI individual of French nationality.

ORGANISM      Homo sapiens
              Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
              Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
              Catarrhini; Homidae; Homo.
              1 (bases 1 to 387)
              Murray,J., Sheffield,V., Weber,J.L., Dux,G. and Buetow,K.H.
              Cooperative Human Linkage Center
              Unpublished (1995)
              Synonyms: GATA47F01, CHLC.GATA47F01.T15166
              COMMENT
              GDB: G00-365-330

```

```

Contact: Dr. Jeffrey C. Murray
Dori
The University of Iowa
Department of Pediatrics, Iowa City, IA 52242, USA
Tel: (319) 356-3508
Fax: (319) 356-3347
Email: jelf-murray@uiowa.edu

Primer A: TATGAGCTAATGCCTTACAGTAA
Primer B: CAGGAAATAGACACACAGC
STS size: 117

PCR Profile:
denature: 30 seconds at 94 degrees C
annealing: 75 seconds at 55 degrees C
extension: 15 seconds at 72 degrees C
PCR cycles: 27
extension: 6 minutes at 72 degrees C

Protocol:
Template: 30ng genomic DNA
Primer: each 1.5 pmole
dNTPs: each 200 uM
Tag Polymerase: 0.3 units
Total Vol: 10 uL

Buffer:
MgCl2: 1.5mM
KCl: 50mM
Tris: 10mM
pH: 8.3

Equivalent set: GATA47G01

FEATURES
source      Location/Qualifiers
            1..387
            /organism="Homo sapiens"
            STS      233..349
            primer_bind 233..256
            primer_bind complement(329..349)
BASE COUNT 99 a 76 c 74 g 126 t 12 others
ORIGIN

Query Match      82.0%; Score 16.4; DB 14; Length 387;
Best Local Similarity 94.4%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 TGCTGTAATCTAATGC 18
        |||||||||||||||
Db 119 TGCTGTAATCTAATCC 102

```

Search completed: September 13, 1999, 15:55:35
 Job time: 4563 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:31 ; Search time 255.05 seconds
(without alignments)
19.619 Million cell updates/sec

Title: US-09-325-095-21

Perfect score: 20
Sequence: 1 TGGCTGAATTCCTAATGCTG 20

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	20	100.0	20	059020	Antisense PCR prim
2	20	100.0	3498	057012	Ptdins 3-kinase 11
3	16.4	82.0	110000	X20248_01	Continuation (2 of
4	15.8	79.0	1797	004597	Equine hemagglutinin
5	15.8	79.0	1788	N71067	Sequence encoding
6	15.8	79.0	1762	029111	EYV HA (A1/Fontain
7	15.8	79.0	1762	029112	EYV HA (A2/Suffolk
8	15.8	79.0	1698	V49391	EYV Fontainebleau
9	15.8	79.0	8395	X13154	Enterococcus faeca
10	15.4	77.0	53585	X20251	Borrelia burgdorfe
11	15.2	76.0	226	020447	B.burgdorferi str
12	15.2	76.0	226	020448	B.burgdorferi str
13	15.2	76.0	2976	T04687	Black widow spider
14	15.2	76.0	3706	T04688	Black widow spider
15	15.2	76.0	4278	V52286	Streptococcus pneu
16	15.2	76.0	110000	V21209_15	Continuation (16 o
17	15.2	76.0	5910	V66798	Rice bacterial lea
18	15.2	76.0	13856	V74342	Staphylococcus aur
19	15.2	76.0	17764	X13238	Enterococcus faeca
20	15.2	76.0	31517	X13117	Enterococcus faeca
21	15.2	76.0	1014	X20074	Enterococcus faeca
22	15.2	76.0	892	X20075	Enterococcus faeca
23	15.2	76.0	110000	X20248_01	Continuation (2 of
24	15.2	76.0	21170	X20535	Polynucleotide seq
25	14.8	74.0	1189	071003	Clone V51 hybridiz
26	14.8	74.0	1512	V39970	Saccharomyces cere
27	14.8	74.0	3930	V39971	Saccharomyces cere
28	14.8	74.0	21338	V52153	Streptococcus pneu
29	14.8	74.0	1267	V65201	DNA encoding a S.
30	14.8	74.0	6146	X13076	Enterococcus faeca
31	14.8	74.0	2639	X52226	Protein PRO220 CDN
32	14.4	72.0	350	T26581	Human gene signatu
33	14.4	72.0	2558	T59360	Mouse Fas-associat
34	14.4	72.0	3833	V17116	Metabotropic gluta
35	14.4	72.0	3321	V04206	Human metabotropic
36	14.4	72.0	3321	V04207	Human metabotropic
37	14.4	72.0	3567	X13164	Enterococcus faeca
38	14.4	72.0	251	X12586	Human diallelic po
39	14.4	72.0	110000	X20248_03	Continuation (4 of
40	14.2	71.0	761	N80472	Sequence specific
41	14.2	71.0	1052	N80475	Genomic form (7P81
42	14.2	71.0	1052	N82441	Genomic form (7P87
43	14.2	71.0	1052	N82442	Genomic form (7P87

ALIGNMENTS

c 44 14.2 71.0 8937 1 020602 Nfi gene, DNA sequ
45 14.2 71.0 1845 1 022614 R. Communis syntha

RESULT 1
ID 059020 standard; DNA; 20 BP.
AC 059020:

DE 12-APR-1994 (first entry)
DE Antisense PCR primer for p110.
KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ss; amplification; p110.

OS Synthetic.
PN W09321328-A.
PD 28-OCT-1993.

PE 13-APR-1993; G00761.
PR 13-APR-1992; GB-008135.

PA (LUDW-) LUDWIG INST CANCER RES.
PI Dhond R, Fry MT, Gout I, Hiles ID, Otsu M, Panayotou G;

PI Parker PJ, Volinia S, Waterfield MD;
PI WPI; 93-351738/44.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity; useful for controlling cell proliferation

PS Example 1; Page 41; 146pp; English.
CC An SGRAP-1 cell line was established by transfection of bovine
CC adrenal cortex zona fasciculata cells with pSVneo. For p110

CC poly (A) + RNA was reversed transcribed with rth DNA polymerase and the
CC antisense primer shown. For DNA polymerisation a sense primer was added
CC and PCR performed. The prod. was p110.

CC See also 051155-6, 059012-23 and 057522-3.
SQ Sequence 20 BP: 5 A; 3 C; 4 G; 8 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCTGAATTCCTAATGCTG 20
Db 1 TGGCTGAATTCCTAATGCTG 20

RESULT 2
ID 057012/c
ID 057012 standard; cDNA to mRNA; 3498 BP.

AC 057012;
DE 31-AUG-1994 (first entry)
DE Ptdins 3-kinase 110 kD catalytic subunit cDNA.
KW 110 kD catalytic subunit; phosphatidyl inositol 3-kinase;
KW transformation; Schizosaccharomyces pombe; nmt promoter; chlamyde;
KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
KW blood vessel plaques; ss.

OS Bos taurus.
FH Key Location/Qualifiers
FT cds 1..3207
FT /tag- a
FT /product- p110

PN W09403609-A.
PD 17-FEB-1994.
PD 05-AUG-1993; G01651.

PE 05-AUG-1992; GB-016654.
PR (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
PI WPI; 94-065697/08.
DR P-PSDB; R46294.

PT Eukaryotic cells transformed with mammalian phospholipid or
PT protein kinase DNA - useful in assays for compounds involved in
PT cell growth regulation and for treating cancers

```
PS Disclosure: Fig 1; 71pp; English.  
CC This sequence encodes the 110 kD catalytic subunit of the phosphatidy  
CC inositol (PtdIns) 3-kinase. This sequence was transformed into  
CC Schizosaccharomyces pombe cells under the regulatory control of the  
CC nmt promoter in an embodiment of the invention. In the presence of  
CC thiamine the promoter is inactive and the cells carrying the Ptdins  
CC catalytic subunit plasmid grow as the parental strain. In the absence  
CC of thiamine the nmt promoter functions and the Ptdins 3-kinase  
CC catalytic subunit is induced. Ptdins activity is substantially  
CC increased under these conditions. Cells containing constructs such as  
CC this, are useful in assays for detecting compounds involved in cell  
CC growth regulation. It is also used as the basis for detecting  
CC compounds for treating cancers and the formation of blood vessel  
CC plaques.
```

```
SO Sequence      3498 BP;       1148 A;        627 C;       721 G;       1002 T;
```

```
Query Match  
Best Local Similarity 100.0%; Score 20; DB 1; Length 3498;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 TGCTGTAATTCATAAGCTG 20  
          |||  
DB      3325 TGCTGTAATTCATAAGCTG 3306
```

```
DESCRIPTION      3  
Continuation (2 of 10) of X20248 from base 100001 (Borrelia burgdorferi polynucleotide s  
MP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
MP Fragment Name      Begin      End  
MP X20248_00           1         110000  
MP X20248_01          10001        210000  
MP X20248_02          20001        310000  
MP X20248_03          30001        410000  
MP X20248_04          40001        510000  
MP X20248_05          50001        610000  
MP X20248_06          60001        710000  
MP X20248_07          70001        810000  
MP X20248_08          80001        910000  
MP X20248_09          90001       910715
```

```
Query Match      82.0%; Score 16.4; DB 1; Length 110000;  
Best Local Similarity 94.4%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      3 CTGTAAATTCATAGCTG 20  
          |||  
DB      30454 CTGTAAATTCATAGCTG 30471
```

```
RESULT      4  
004597/c  
ID   004597 standard; DNA; 1797 BP.  
DC   004597;  
DT   10-MAR-1993 (revised)  
DT   02-OCT-1990 (first entry)  
DE   Equine haemagglutinin H3 (Eiv-A2).  
KW   Recombinant vaccines; equine Influenza virus; haemagglutinin; H3;  
KW   neuraminidase; NB; ss.  
OS   Equine Influenza Virus.  
FH   Key Location/Qualifiers  
FT   cds          54..1748  
FT               /*tag= a  
FT               /product=EIV-A2 haemagglutinin H3  
FT               102..1182  
FT               /*tag= b  
FT               /product=N-terminal HA1 50KD portion  
FT               1186..1748  
FT               /*tag= c  
FT               /product=C-terminal HA2 27KD portion  
FN   US4920213-A.
```

```

PD 24-APR-1990.
PF 21-JUN-1986; 888250.
PR 20-JUN-1985; US-747020.
PR 21-JUL-1986; US-888250.
PA (BIOT-) Biotech Res Partner.
PI Dale B, Cordell B;
DR WPI: 90-163647/21.
P-PSDB: R04943.
PT Recombinant vaccines against equine influenza virus - produced using DNA
PT sequences encoding haemagglutinin and neuraminidase glycoprotein(s).
PS Disclosure: 27pp; English.
CC There are nine neuraminidase (NA) subtypes and twelve haemagglutinin (HA)
CC subtypes. The strain carrying H3N8 glycoproteins is designated equine
CC influenza virus (EIV)-A2. The cDNA sequences encoding these
CC will be useful in the construction of diagnostic probes for the
CC disease and of probes for obtaining new cDNAs of the mutated form of
CC the virus. Recombinant vaccines are produced.
CC See also 004596-004599.
SC Sequence 1797 BP; 639 A; 341 C; 413 G; 404 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1797;
Best Local Similarity 89.5%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0

Ox 1 TGCTGTAATTCCTAATGCT 19
      ||||| |||||
Db 464 TGCTGTGAATTCCTAATGCT 446

RESULT 5
N71067/C
ID N71067 standard; DNA; 1788 BP.
AC N71067;
DE 08-MAR-1991 (first entry)
DI Sequence encoding equine influenza virus strain H3N8 (EIV-A2)
DE haemagglutinin protein H3.
DE HA; vaccine; Vaccine; ds.
OS Equine influenza virus.
KS Key Location/Qualifiers
FT cds 54..1748
FT mat_peptide 102..1082
FT /*tag- a
FT /*tag- b
FT /*label- HA 1
FT 1086..1748
FT /*tag- c
FT /*label- HA 2

W08607593-A.
PD 31-DEC-1986.
PF 20-JUN-1986; U01343.
PR 20-JUN-1985; US-747020.
PA (BIOT-) BIOTECHN RES PARTNER.
PI Dale B, Cordell B;
DR WPI: 87-007191/01.
P-PSDB: E70711.
PT Preventing equine influenza virus infection - using recombinant
PT vaccines produced using DNA sequences encoding haemagglutinin and
PT neuraminidase glyco:proteins
PS Disclosure: Fig 2; 63pp; English.
CC Peptides derived from the haemagglutinin H7 and H3 and neuraminidase
CC N7 and N8 genes may be used to derive antigenic peptides useful in
CC vaccination against equine influenza virus infection.
CC Abs raised to the peptides may be used in diagnosis of the infection
CC and construction of probes to mutated forms of the virus.
SC Sequence 1788 BP; 637 A; 340 C; 409 G; 402 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1788;
Best Local Similarity 89.5%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0

Ox 1 TGCTGTAATTCCTAATGCT 19
      ||||| |||||

```

Db 464 TGCCTGATCTATGTT 446

RESULT 6

ID 029111/c 029111 standard; DNA: 1762 BP.

AC 029111; 24-FEB-1993 (first entry)

DE EIV HA (A2/Fontainebleau/79).

KW Equine influenza virus; EIV; haemagglutinin; HA: A1/Fontainebleau/79;

KW expression cassette; NYVAC; ALVAC; recombinant vector;

KW polymerase chain reaction; PCR; vaccinia virus; H6 promoter;

KW canarypox virus; Copenhagen vaccine strain; virulence factor;

KW deletion loci; recipient loci; ss.

OS Synthetic.

PN WO9215672-A.

PD 17-SEP-1992.

PF 09-MAR-1992; 001906.

PR 07-MAR-1991; US-666056.

PR 11-JUN-1991; US-713967.

PR 06-MAR-1992; US-847951.

PA (VIR-) VIROGENETICS CORP.

PI Cox WI, De Taisne C, Francis J, Gettig RR, Johnson GP,

PI Limbach KJ, Norton EK, Paolletti E, Perkins ME, Pincus SE,

PI Riviere M, Tartaglia J, Taylor J;

PI WPI: 92-331718/40.

PT Vaccine comprises recombinant, attenuated pox-virus - use for

PT vaccinating against viral infections such as rabies, hepatitis B,

PT HIV, HSV, EBV, CMV, mumps etc.

PS Disclosure; Fig 24; 456pp; English.

CC The sequence given encodes the equine influenza virus (EIV)

CC haemagglutinin (HA) (A1/Fontainebleau/79). This sequence was used to

CC generate an expression cassette for the insertion of the EIV HA

CC gene into NYVAC and ALVAC recombinant vectors. The HA gene sequence

CC was isolated from an EIV cDNA library and was amplified by polymerase

CC chain reaction. The HA gene sequence was fragmented and then

CC reconstituted aligned with the vaccinia virus H6 promoter. NYVAC is

CC derived from a Copenhagen vaccine strain of vaccinia virus and ALVAC

CC is derived from a canarypox virus which has been modified by deletion

CC of non-essential regions of the genome encoding known or potential

CC virulence factors. The deletion loci of both vectors were engineered

CC as recipient loci for the insertion of foreign genes. See also

CC Q35501-864.

SQ Sequence 1762 BP; 626 A; 331 C; 395 G; 410 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1762;

Best Local Similarity 89.5%; Pred. No. 57;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCCTGATCTATGTT 19

Db 440 TGCCTGATCTATGTT 422

029112/c

ID 029112 standard; DNA: 1762 BP.

AC 029112; 24-FEB-1993 (first entry)

DE EIV HA (A2/Suffolk/89).

KW Equine influenza virus; EIV; haemagglutinin; HA: A2/Suffolk/89;

KW expression cassette; NYVAC; ALVAC; recombinant vector; M13;

KW polymerase chain reaction; PCR; vaccinia virus; 13L promoter;

KW canarypox virus; Copenhagen vaccine strain; virulence factor;

KW deletion loci; recipient loci; ss.

OS Synthetic.

PN WO9215672-A.

PD 17-SEP-1992.

PF 09-MAR-1992; 001906.

PR 07-MAR-1991; US-666056.

PR 11-JUN-1991; US-713967.

PR 06-MAR-1992; US-847951.

PA (VIR-) VIROGENETICS CORP.

PI Cox WI, De Taisne C, Francis J, Gettig RR, Johnson GP,

PI Limbach KJ, Norton EK, Paolletti E, Perkins ME, Pincus SE,

PI Riviere M, Tartaglia J, Taylor J;

PI WPI: 92-331718/40.

PT Vaccine comprises recombinant, attenuated pox-virus - use for

PT vaccinating against viral infections such as rabies, hepatitis B,

PT HIV, HSV, EBV, CMV, mumps etc.

PS Disclosure; Fig 25; 456pp; English.

CC The sequence given encodes the equine influenza virus (EIV)

CC haemagglutinin (HA) (A2/Suffolk/89). This sequence was used to

CC generate an expression cassette for the insertion of the EIV HA

CC gene into NYVAC and ALVAC recombinant vectors. The HA gene sequence

CC was isolated by polymerase chain reaction from an M13 clone. Some

CC non-conserved base-changes were observed due to the amplification

CC process. The EIV HA gene was linked to the 13L promoter and inserted

CC into a vaccinia insertion plasmid deleted for ORF's C6L-K1L. NYVAC is

CC derived from a Copenhagen vaccine strain of vaccinia virus and ALVAC

CC is derived from a canarypox virus which has been modified by deletion

CC of non-essential regions of the genome encoding known or potential

CC virulence factors. The deletion loci of both vectors were engineered

CC as recipient loci for the insertion of foreign genes. See also

CC Q35501-864.

SQ Sequence 1762 BP; 637 A; 328 C; 384 G; 412 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1762;

Best Local Similarity 89.5%; Pred. No. 57;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCCTGATCTATGTT 19

Db 440 TGCCTGATCTATGTT 422

029112/c

ID 029112 standard; DNA: 1698 BP.

AC 029112; 28-OCT-1998 (first entry)

DE EIV Fontainebleau strain haemagglutinin gene.

KW Multivalent vaccine; horse; pathogen; respiratory disease; EHV; EIV;

KW Clostridium tetani; Borrelia burgdorferi; equine influenza virus; EEEV;

KW Eastern equine encephalomyelitis virus; equine herpesvirus; WEEV; VEEV;

KW Western equine encephalomyelitis virus; digestive disease; rabies virus;

KW Venezuelan equine encephalomyelitis virus; vector; primer; PCR.

OS Equine influenza virus.

FT Key Location/Qualifiers

FT CDS 1..1698

FT /tag- a

FR2751226-A1. /product- "haemagglutinin protein"

PN 23-JAN-1998.

PD 19-JUL-1996; 009400.

PF 19-JUL-1996; FR-009400.

PA (INMR) RHONE MERIEUX SA.

PI Audonnet JCF, Bouchardon A, Riviere MEA.

DR WPI: 98-112826/11.

DR P-PSDB: W44946.

PT Multi-valent polynucleotide vaccines against equine pathogens

PT consist of at least 3 plasmids able to express protective antigens

PT from specified viruses

PS Example 14; Fig 8; 49pp; French.

CC The invention relates to a multivalent vaccine for protecting horses

CC against several pathogens, especially pathogens associated with

CC respiratory and digestive diseases. The pathogens are especially

CC selected from equine herpesvirus (EHV), equine influenza virus (EIV),

CC Clostridium tetani, Borrelia burgdorferi, Eastern, Western or Venezuelan

CC equine encephalomyelitis viruses (EEEV, WEEV and VEEV, respectively) and

CC rabies virus. The vaccines are preferably composed of polynucleotide

CC sequences encoding 3 antigens, all as part of vectors.

CC This sequence represents the coding region of the ETV Fontainebleau
 CC strain haemagglutinin gene. The sequence was subcloned into the plasmid
 CC pVR1012 to generate plasmid PAB099 for use in the vaccine.
 CC Sequence 1698 BP; 604 A; 319 C; 384 G; 391 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1698;
 Best Local Similarity 89.5%; Pred. No. 57;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCT 19
 ||||| ||||| ||||| ||
 DB 411 TGCTGTAATTCATGCT 393

RESULT 9
 ID X13154 standard; DNA: 8395 BP.
 AC X13154;
 DT 19-MAR-1999 (first entry)
 DE Enterococcus faecalis genome contig SEQ ID NO:217.
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;
 KW vaccine; attenuation; computer readable medium; ds.
 OS Enterococcus faecalis.
 PN M09850555-42.
 PD 12-NOV-1998.
 PE 04-MAY-1998; U08985.
 PR 14-NOV-1997; US-066009.
 PR 06-MAY-1997; US-044031.
 PR 16-MAY-1997; US-046655.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Dillon PJ, Kunsch CA;
 DR WPI; 99-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.

PS Claim 1: Page 1103-1107; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC X13938 to X13919 represent these nucleotide sequences which are primary
 CC nucleotide sequences, also known as contigs. The computer-based system
 CC can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism. In vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.
 CC Sequence 8395 BP; 2616 A; 1692 C; 1478 G; 2607 T;

Query Match 79.0%; Score 15.8; DB 1; Length 8395;
 Best Local Similarity 89.5%; Pred. No. 63;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCT 19
 ||||| ||||| ||||| ||
 DB 6161 TGCTGTAATTCATGCT 6179

RESULT 10
 X20251/c
 ID X20251 standard; DNA: 53585 BP.
 AC X20251;
 DT 04-MAY-1999 (first entry)

DE Borrelia burgdorferi polynucleotide sequence #4.
 CC Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.

OS Borrelia burgdorferi.
 PN W09858943-A1.
 PD 30-DEC-1998.
 PE 18-JUN-1998; U12764.
 PR 03-SEP-1997; US-057483.
 PR 20-JUN-1997; US-050359.
 PR 22-JUL-1997; US-053844.
 PR 22-JUL-1997; US-053377.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI (MED-) MEDIMUNE INC.
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
 PI White OS;
 DR WPI; 99-081217/07.

PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 PS Claim 1: Page 801-831; 1128pp; English.
 CC X20248 to X20402 represent polynucleotide sequences isolated from
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.

CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 CC Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T;

Query Match 77.0%; Score 15.4; DB 1; Length 53585;
 Best Local Similarity 94.1%; Pred. No. 11e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCT 17
 ||||| ||||| ||||| ||
 DB 11703 TGCTGTAATTCATGCT 11687

RESULT 11

ID Q20447 standard; DNA: 226 BP.
 AC Q20447;

DT 15-APR-1992 (first entry)
 DE B.burgdorferi strain P/B1 flagellin gene fragment.

KW tick-borne spirochete; Lyme disease; polymerase chain reaction; PCR;
 KW primer region; G1; G2; VS3; ss.

OS Borrelia burgdorferi.
 PN W09119814-A.
 PD 26-DEC-1991.

PE 12-JUN-1991; U04190.
 PR 15-JUN-1990; US-538957.
 PR 25-APR-1991; US-691188.

PA (BAXT) BAXTER DIAGNOSTICS.
 PI Picken RN, Ammons HC;
 DR WPI; 92-024430/03.

PT New DNA primers useful in diagnosis of Lyme disease - comprise
 PT complementary strand specifically flanking Borrelia burgdorferi
 PT sequence

PS Disclosure: Fig 6; 47pp; English.
 CC This sequence corresponds to nucleotides 517 to 742 of the
 CC B.burgdorferi flagellin gene sequence of Q20720. The region was
 CC amplified using primers flanking a region of relative non-homology
 CC between B.burgdorferi and B.hemissi flagellin gene sequences. A
 CC comparison of amplified sequences from a number of B.burgdorferi
 CC strains (see Q20446-9) indicated several positions of mismatched
 CC bases. By selecting a region containing minimally 5 mismatched
 CC bases (i.e. between nucleotides 78 to 126), three probes were
 CC obtained (see Q20720, Q21925-6) which could distinguish three
 CC subgroups of B.burgdorferi in an assay in which the interprimer
 CC region was first amplified and then probed. Other strains which
 CC belong to the P/B1 subgroup are G1, G2, VS3 (all with sequences
 CC identical to P/B1) and VS185. See also Q20721, Q20723-5 and Q21927.

CC Sequence 226 BP; 65 A; 46 C; 53 G; 62 T;

Query Match 76.0%; Score 15.2; DB 1; Length 226;
Best Local Similarity 85.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCG 20
|||||
DB 39 TGCTGTAATTCATGCG 58

RESULT 12
020448
ID Q20448 standard; DNA: 226 BP.
AC Q20448;
DT 15-APR-1992 (first entry)
DE B.burgdorferi strain VS185 flagellin gene fragment.
KW tick-borne spirochaete; Lyme disease; polymerase chain reaction; PCR;
KM Interprimer region; P/B1 subgroup; G1; G2; VS3; SS.
OS Borrelia burgdorferi.
PN W09119814-A.
PD 26-DEC-1991.
PF 12-JUN-1991; U04190.
PR 15-JUN-1990; US-538957.
PR 25-APR-1991; US-691188.
PA (BAXT) BAXTER DIAGNOSTICS.
PI Picken RN, Ammons HC.
PT WPI: 92-024430/03.
PT New DNA primers useful in diagnosis of Lyme disease - comprise complementary strand specificity flanking Borrelia burgdorferi PT sequence
PS Disclosure: Fig 6; 47pp; English.
CC This sequence corresponds to nucleotides 517 to 742 of the CC B.burgdorferi flagellin gene sequence of Q20720. The region was CC amplified using primers flanking a region of relative non-homology CC between B.burgdorferi and B.hemissi flagellin gene sequences. A CC comparison of amplified sequences from a number of B.burgdorferi CC strains (see Q20446-9) indicated several positions of mismatched CC bases (i.e. between nucleotides 78 to 126); three probes were CC obtained (see Q20720, Q21925-6) which could distinguish three CC subgroups of B.burgdorferi in an assay in which the interprimer ~~~~~ was first amplified and then probed. Strain VS185 belongs to CC the P/B1 subgroup, along with strains G11, G2, VS3.
CC See also Q20721, Q20723-5 and Q21927.
CC Sequence 226 BP; 67 A; 48 C; 52 G; 59 T;

Query Match 76.0%; Score 15.2; DB 1; Length 226;
Best Local Similarity 85.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCG 20
|||||
DB 39 TGCTGTAATTCATGCG 58

RESULT 13
T04687
ID T04687 standard; DNA: 2376 BP.
AC T04687;
DT 16-MAR-1996 (first entry)
DE Black widow spider delta-latroinsectotoxin truncated gene DNA.
KW Black widow spider; insecticide; pesticide; delta-latroinsectotoxin;
KM spider venom; neurotoxin; toxin; ds.
OS Latrodectus mactans tenebrimontatus.
FH Key Location/Qualifiers
FT cds 1..2976 /*tag= a
PN GB2288807-A.
PD 01-NOV-1995.
PF 24-APR-1994; 008298.
PR 27-APR-1994; GB-008466.
PR 27-APR-1994; GB-008466.

PA (BRTE-) BRITISH TECHNOLOGY GROUP LTD.
PI Bell DR, Dulubova I, Galkina TG, Grishin E, Khovtchey MV;
PI Krasnoperov V, Pluzhnikov KA, Shamotlenko OG, Usherwood PMR;
PI Volkova T, Galkina T, Khovtchey MV;
DR WPI: 95-360758/47.
DR P-PSDB: R80096.
PT Polypeptide(s) expressed by truncated genes, esp. spider
PT delta-latroinsectotoxin - also related non toxic precursor
PS Claim 10; Page 31-35; 62pp; English.
CC This truncated gene may be expressed recombinantly in E. coli BL21 (CC DE3) cells transformed with pT7-7 vectors comprising the truncated CC form of the sequence. The encoded protein is an insect-specific CC neurotoxin, delta-latroinsectotoxin, which is useful as an oral or CC topical insecticide. Expressing the truncated gene allows large- CC scale production of active toxin in bacteria, eliminating the need CC to extract it from spider venom. This neurotoxin is harmless to CC mammals and plants.
SQ Sequence 2376 BP; 1089 A; 455 C; 571 G; 861 T;

Query Match 76.0%; Score 15.2; DB 1; Length 2976;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCG 20
|||||
DB 896 TGATGTAATTCATGCG 915

RESULT 14
T04688
ID T04688 standard; DNA: 3706 BP.
AC T04688;
DT 16-MAR-1996 (first entry)
DE Black widow spider delta-latroinsectotoxin gene precursor DNA.
KW Black widow spider; insecticide; pesticide; delta-latroinsectotoxin;
KM spider venom; neurotoxin; toxin; ds.
OS Latrodectus mactans tenebrimontatus.
FH Key Location/Qualifiers
FT cds 45..3686 /*tag= a
PN GB2288807-A.
PD 01-NOV-1995.
PF 24-APR-1994; 008298.
PR 27-APR-1994; GB-008466.
PA (BRTE-) BRITISH TECHNOLOGY GROUP LTD.
PI Bell DR, Dulubova I, Galkina TG, Grishin E, Khovtchey MV;
PI Krasnoperov V, Pluzhnikov KA, Shamotlenko OG, Usherwood PMR;
PI Volkova T, Galkina T, Khovtchey MV;
DR WPI: 95-360758/47.
DR P-PSDB: R80097.
PT Polypeptide(s) expressed by truncated genes, esp. spider
PT delta-latroinsectotoxin - also related non toxic precursor
PS Claim 62; Page 38-43; 62pp; English.
CC This gene encodes an insect-specific neurotoxin, delta- CC lactroinsectotoxin, which is useful as an oral or topical CC insecticide. The gene has been cloned into E. coli (HMS 174/ CC PT7.deltatl). This neurotoxin is harmless to mammals and plants.
SQ Sequence 3706 BP; 1342 A; 588 C; 692 G; 1084 T;

Query Match 76.0%; Score 15.2; DB 1; Length 3706;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCG 20
|||||
DB 896 TGATGTAATTCATGCG 915

RESULT 15

V52286/C
 ID V52286 standard; DNA: 4278 BP.
 AC V52286;
 DT 23-OCT-1998 (first entry)
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:153.
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KM computer readable medium; vaccine; pharmaceutical composition; ds.
 OS Streptococcus pneumoniae.
 PN W09818931-R2.
 PD 07-MAY-1998.
 PF 30-OCT-1997; U19588.
 PR 31-OCT-1996; US-029960.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
 PI Kunsch CA, Rosen CA;
 DR WPI: 98-272225/24.
 PT Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae
 PS Claim 1: Page 1018-1020; 1409pp; English.
 CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
 CC on it, or a representative fragment or a sequence at least 95% identical
 CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
 CC pneumoniae. The present invention also describes an isolated nucleic acid
 CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
 CC by a process comprising: (a) screening a genomic DNA library using as a
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
 CC to 391, identifying members of the library which contain sequences
 CC that hybridize to the target sequence and isolating the nucleic acid
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
 CC from an organism, amplifying nucleic acid molecules whose nucleotide
 CC sequence is homologous to amplification primers derived from the
 CC fragment of the S. pneumoniae genome to prime the amplification and
 CC isolating the amplified sequences. The computer readable medium can be
 CC used in a computer-based system for identifying fragments of the
 CC S. pneumoniae genome of commercial importance, or expression modulating
 CC fragments of the S. pneumoniae genome. Products from the present
 CC invention can be used in diagnosis kits and assays, and pharmaceutical
 CC compositions and vaccines for S. pneumoniae.
 SQ Sequence 4278 BP; 1209 A; 883 C; 947 G; 1239 T;

Query Match 76.0%; Score 15.2; DB 1; Length 4278;
 Best Local Similarity 85.0%; Pred. No. 1.2e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCTGTAATTCATGCTG 20
 |||||
 DB 633 TGCTGTAATTCATGCTG 614

Search completed: September 13, 1999, 15:59:33
 Job time: 4800 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

nm nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:56 ; Search time 1694.61 Seconds

(without alignments)
23.280 Million cell updates/sec

Title: US-09-325-095-21

Perfect score: 20
Sequence: 1 TGCCTGAATCTAATGCTG 20

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query	Length	DB	ID	Description
1	20	100.0	475	25	N44819	N44819 YY39a10.r1
2	20	100.0	343	31	AA298585	AA298585 EST114216
3	20	100.0	633	49	A1632172	A1632172 t585907.x
4	18.4	92.0	136	25	N84360	N84360 KR7998F Hum
5	18.4	92.0	340	34	AA508979	AA508979 MBAPCWLHO
6	17.4	87.0	246	21	R10988	R10988 YF38C09.r1
7	16.8	84.0	580	36	C80053	C80053 C80053 Mous
8	16.8	84.0	208	43	A1102178	A1102178 EST211467
9	15	80.0	613	39	C83915	C83915 C83915 Dict
10	15	80.0	580	44	AU037810	AU037810 AU037810
11	15.8	79.0	496	22	R52778	R52778 Y996C11.r1
12	15.8	79.0	500	22	R59012	R59012 Y996C01.r1
13	15.8	79.0	214	25	N50767	N50767 YY90F02.r1
14	15.8	79.0	642	27	AA020934	AA020934 z664f12.r
15	15.8	79.0	305	28	AA120789	AA120789 zK00611.r
16	15.8	79.0	433	30	AA248994	AA248994 JA00A091.
17	15.8	79.0	555	32	AA329331	AA329331 EST33605
18	15.8	79.0	450	33	AA417468	AA417468 m1/d01 YE
19	15.8	79.0	453	33	AA433124	AA433124 JA00A201.
20	15.8	79.0	514	48	A1561906	A1561906 Y183a09.x
21	15.8	79.0	337	50	AV047614	AV047614 AV047614
22	15.4	77.0	384	24	H72799	H72799 YU07410.r1
23	15.4	77.0	334	27	AA050566	AA050566 m16a12.r
24	15.4	77.0	308	27	W91533	W91533 m184C11.r1
25	15.4	77.0	554	34	AA524662	AA524662 nh35e08.s
26	15.4	77.0	310	39	AA831652	AA831652 oc82h10.s
27	15.4	77.0	380	46	A1443743	A1443743 sA45a06.y
28	15.4	77.0	337	49	A1625843	A1625843 ty65d06.x
29	15.2	76.0	307	20	T05994	T05994 EST03883 FE
30	15.2	76.0	494	20	T58174	T58174 YB26F04.r1
31	15.2	76.0	322	20	Z42752	Z42752 HSC08H081 n
32	15.2	76.0	322	21	F00595	F00595 HSC31H051 n
33	15.2	76.0	314	21	F13112	F13112 HSC31H051 n
34	15.2	76.0	363	21	R09683	R09683 Y123h05.r1
35	15.2	76.0	342	21	T75528	T75528 Y63d04.r1
36	15.2	76.0	349	21	T82265	T82265 Yd44g03.s1
37	15.2	76.0	381	22	R50603	R50603 Y160g08.r1
38	15.2	76.0	417	22	R52980	R52980 Y85e10.s1
39	15.2	76.0	472	22	R53776	R53776 Y102e11.r1
40	15.2	76.0	426	22	R55507	R55507 Y175g11.r1
41	15.2	76.0	445	22	R73800	R73800 Y197g09.r1
42	15.2	76.0	456	22	R73843	R73843 Y155f11.r1
43	15.2	76.0	464	23	H44728	H44728 YP24f05.r1
44	15.2	76.0	444	23	H45554	H45554 Y072d06.r1
45	15.2	76.0	510	51	A1708406	A1708406 at13d04.x

ALIGNMENTS

RESULT 1
N44819/c N44819 475 bp mRNA EST 13-FEB-1996
LOCUS YY39a10.r1 Soares melanocyte 2NbhM Homo sapiens
DEFINITION IMAGE:273594 5' similar to SW:P1A_BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
ACCESSION N44819
N44819

NID G1185985
 VERSION N44819.1 GI:1185985
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 LITERATURE
 1 (bases 1 to 475)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 The Washu-Merck EST Project
 Unpublished (1995)
 On May 8, 1995 this sequence version replaced gi:1800198.
 COMMENT
 TITLE The Washu-Merck EST Project
 JOURNAL Unpublished (1995)
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: T7
 High quality sequence stop: 402.
 Location/Qualifiers
 1..475
 /organism="Homo sapiens"
 /db_xref="GDB:3883236"
 /db_xref="taxon:9606"
 /clone_image="273594"
 /clone_lib="Soares melanocyte 2NBM"
 /sex="Male"
 /tissue_type="melanocyte"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTCACATCTGACGTGGAGCGCGCGAGTATTTTATTTTATTTT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library constructed by Bento Soares and
 M. Fatima Bonaldi. RNA from normal foreskin melanocytes
 (FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 170 a 90 c 85 g 128 t 2 others
 ORIGIN
 Query Match 100.0%; Score 20; DB 25; Length 475;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGCTGTAATTCATATGCTG 20
 Db 372 TGCTGTAATTCATATGCTG 353
 RESULT 2
 AA298585 343 bp mRNA EST 18-APR-1997
 LOCUS EST114216 HSC172 cells II Homo sapiens cDNA 5' end similar to
 DEFINITION phosphatidylinositol 3-kinase, alpha catalytic subunit, mRNA
 sequence.
 accession AA298585
 NID G11950938
 VERSION AA298585.1 GI:1950938
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE
 AUTHORS
 1 (bases 1 to 343)
 Adams, M.D., Kerlavage, A.R., Flieschmann, R.D., Fulcher, R.A.,
 Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstein, K.G., Gocayne, J.D.,
 White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, Wai, C.,
 Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,
 Fitzgerald, L.M., Fitch, W.M., Fitchman, J.L., Geoghagen, N.S.,
 Glodde, A., Gnehm, C.L., Hanna, W.C., Hedblom, E., Hinkle, P.S., Jr.,
 Kelley, J.M., Kelley, J.C., Liu, L.-I., Marnaros, S.M., Merrick, J.M.,
 Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
 Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
 Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
 Bedarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
 Dimke, D., Feng, D.F., Ferris, A., Fischer, C., Hastings, G.A.,
 He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,
 Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H.,
 Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M.,
 Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,
 Fraser, C.M. and Venter, J.C.
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl.), 3-174 (1995)
 96026280
 JOURNAL
 MEDLINE
 COMMENT
 On Sep 12, 1996 this sequence version replaced gi:1392803.
 Other ESTs: TIGR168479
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/db/hgi.html>)
 Seq primer: M3 Reverse.
 Location/Qualifiers
 1..343
 /organism="Homo sapiens"
 /db_xref="ATCC (Inhost):179595"
 /db_xref="taxon:9606"
 /clone_lib="HSC172 cells II"
 /cell_type="fibroblast"
 /cell_line="HSC172 (60PDL)"
 /dev_stage="fetal"
 /note="Organ: lung; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI."

BASE COUNT 132 a 61 c 64 g 84 t 2 others
 ORIGIN
 Query Match 100.0%; Score 20; DB 31; Length 343;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGCTGTAATTCATATGCTG 20
 Db 287 TGCTGTAATTCATATGCTG 268
 RESULT 3
 AI632172 633 bp mRNA EST 26-APR-1999
 LOCUS ts85907.x1 NCI-GC6 Homo sapiens cDNA clone IMAGE:228108 3'
 DEFINITION similar to SW-614 BOVIN P32871 PHOSPHATIDYLIINOSITOL 3-KINASE
 CATALYTIC SUBUNIT, ALPHA ISOFORM ;, mRNA sequence.
 accession AI632172
 NID G4683502
 VERSION AI632172.1 GI:4683502
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE *1 (bases 1 to 633)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 7, 1998 this sequence version replaced gi:3121411.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

FEATURES
 source
 Seq primer: -400P from Gibco
 High quality sequence stop: 446.
 Location/Qualifiers

1..633
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="15q26.1"
 /clone="IMAGE:2238108"
 /clone_1lb="NCI-CGAP-GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; plasmid DNA from the normalized library
 NCI-CGAP-GC4 was prepared, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from a pool of 5,000 clones made
 from the same library (cloneids 1257096-1258631,
 1469064-1470983, and 1475592-1476743). Subtraction by
 Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 169 a 109 c 108 g 246 t 1 others

Query Match 100.0%; Score 20; DB 49; Length 633;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCCTGAATCTTAATGCTG 20
 ||||||||||||||||
 Db 223 TGCCTGAATCTTAATGCTG 239

RESULT 4
 LOCUS N84360/c 136 bp mRNA EST 01-APR-1996
 DEFINITION KR7998F Human fetal heart, Lambda ZAP Express Homo sapiens CDNA
 clone KR7998 5' similar to PHOSPHATIDYLINOSITOL 3-KINASE, mRNA
 sequence.
 ACCESSION N84360
 NID g1259985
 VERSION N84360.1 GI:1259985
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 136)
 AUTHORS Liew, C.C.
 TITLE CDNAs from fetal heart (1996)
 JOURNAL Unpublished (1996)
 COMMENT On May 8, 1995 this sequence version replaced gi:800157.

Contact: Liew CC
 Department of Laboratory Medicine and Pathobiology
 University of Toronto
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
 Tel: 4169788758
 Fax: 4169788560
 Email: liewcc@utcc.utoronto.ca
 Seq primer: GAATTAACCTCACTAAAGCG.
 Location/Qualifiers

FEATURES
 source
 1..136
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="KR7998"
 /clone_1lb="Human fetal heart, Lambda ZAP Express"
 /lab_host="E. coli XL1-blue"
 /note="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2:
 XhoI; mRNA was purified from human fetal hearts (8-10
 weeks). CDNA was synthesized using a XhoI-Oligo dT
 adaptor-primer. EcoRI adaptors were ligated, followed by
 digestion with XhoI, for directional cloning into
 predigested lambda ZAP Express."
 BASE COUNT 52 a 29 c 26 g 29 t

Query Match 92.0%; Score 18.4; DB 25; Length 136;
 Best Local Similarity 95.0%; Pred. No. 70;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCCTGAATCTTAATGCTG 20
 ||||||||||||||||
 Db 124 TGCCTGAATCTTAATGCTG 105

RESULT 5
 LOCUS AA508979 340 bp mRNA EST 08-JUL-1997
 DEFINITION MBACWIH03T3 Brugia malayi adult female cDNA (SAW95MLW-BMAC) Brugia
 malayi CDNA clone AFWCIH03 5', mRNA sequence.
 ACCESSION AA508979
 NID g2246856
 VERSION AA508979.1 GI:2246856
 KEYWORDS EST.
 SOURCE Brugia malayi.
 ORGANISM Brugia malayi.
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
 Filarioidea; Onchocercidae; Brugia.
 REFERENCE 1 (bases 1 to 340)
 AUTHORS Blaxter, M.L., Waterfall, M., Daub, J., Lizotte, M., Baron, L. and
 Jones, S.J.
 TITLE Genes expressed in adult female Brugia malayi
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1395421.

Contact: Blaxter ML
 Institute of Cell, Animal and Population Biology
 University of Edinburgh
 Ashworth labs, King's Buildings, West Mains Road, Edinburgh, EH9
 3JF, UK.
 Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk
 The ABI trace of this sequence can be viewed at
<http://www.sanger.ac.uk/brugia/AFWC/MBACWIH03T3.html>
 Seq primer: T3

FEATURES
 source
 1..340
 /organism="Brugia malayi"
 /db_xref="taxon:6279"
 /map="891D05; 6; 6q22.2-6q23.2; 17q21"
 /clone="AFWCWIH03"
 /clone_1lb="Brugia malayi adult female CDNA

```
QY      1 TCGTCTAAATTCATACTG 20  
        |||||  
Db      68 TCGTCTAAATTCATACTG 49
```

RESULT 8
A102178.

LOCUS A1102178 208 bp mRNA EST 31-JAN-1999
 DEFINITION EST211467 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
 RBRCA22 3' end, mRNA sequence.
 ACCESSION A1102178
 MID 93706986
 VERSION A1102178.1 GI:3706986
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 208)
 Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 JOURNAL On Aug 21, 1998 this sequence version replaced.
 COMMENT

CONTACT: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.
 Location/Qualifiers
 1..208
 /organism="Rattus sp."
 /db_xref="ATCC (inhost):2024467"
 /db_xref="taxon:10118"
 /map="19p12-p13.1"
 /clone="RBRCA22"
 /clone_lib="Normalized rat brain, Bento Soares"
 /note="Organ: brain; Vector: pT7Tpac; Site_1: EcoRI;
 Site_2: NciI"

BASE COUNT 57 a 53 c 34 g 64 t
 ORIGIN

Query Match 84.0%; Score 16.8; DB 43; Length 208;
 Best Local Similarity 90.0%; Pred. No. 3.2e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 TGTCTAAATCTATGCTG 20
 |||||
 Db 32 TGTCTGAATCTATGCTG 51

RESULT 9
 C83915 613 bp mRNA EST 28-APR-1999
 LOCUS C83915 Dictyostelium discoideum SS (H. Urushihara) Dictyostelium
 DEFINITION
 accession C83915
 MID 92706847
 VERSION C83915.1 GI:2706847
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum.
 ORGANISM Dictyostelium discoideum.
 Eukaryota; Dictyostelida; Dictyostelium.
 REFERENCE 1 (bases 1 to 613)
 Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
 Yoshino, R., Mitra, B.N., Pl.M., Sato, T., Takemoto, K., Yasukawa, H.,
 Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
 The Dictyostelium developmental cDNA project: generation and
 analysis of expressed sequence tags from the first-finger stage of
 development
 JOURNAL DNA Res. 5 (6), 335-340 (1998)
 MEDLINE 99156227
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2153552.

CONTACT: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
 Email: d402hu@tsukuba.cc.tsukuba.ac.jp
 PROJECT - 'Dictyostelium discoideum cDNA project in Japan'
 POLYX-NO.
 Location/Qualifiers
 1..613
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SSA593"
 /clone_lib="Dictyostelium discoideum SS (H. Urushihara)"
 /dev_stage="slug"
 BASE COUNT 207 a 65 c 104 g 236 t 1 others
 ORIGIN

Query Match 80.0%; Score 16; DB 39; Length 613;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 4 TGTAAATCTATGCT 19
 |||||
 Db 211 TGTAAATCTATGCT 196

RESULT 10
 A0037810 580 bp mRNA EST 29-MAR-1999
 LOCUS A0037810 Dictyostelium discoideum SS (H. Urushihara) Dictyostelium
 DEFINITION
 accession A0037810
 MID 93984563
 VERSION A0037810.1 GI:3984563
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum.
 ORGANISM Dictyostelium discoideum.
 Eukaryota; Dictyostelida; Dictyostelium.
 REFERENCE 1 (bases 1 to 580)
 Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
 Yoshino, R., Mitra, B.N., Pl.M., Sato, T., Takemoto, K., Yasukawa, H.,
 Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
 The Dictyostelium developmental cDNA project: generation and
 analysis of expressed sequence tags from the first-finger stage of
 development
 JOURNAL DNA Res. 5 (6), 335-340 (1998)
 MEDLINE 99156227
 COMMENT On Jan 5, 1998 this sequence version replaced gi:2747316.

CONTACT: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
 Email: d402hu@tsukuba.cc.tsukuba.ac.jp
 PROJECT - 'Dictyostelium discoideum cDNA project in Japan'.
 Location/Qualifiers
 1..580
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SSE323"
 /clone_lib="Dictyostelium discoideum SS (H. Urushihara)"
 /dev_stage="slug"

BASE COUNT 217 a 53 c 86 g 224 t
 ORIGIN
 Query Match 80.0%; Score 16; DB 44; Length 580;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TGTAAATCTAATGCT 19
 DB 89 TGTAAATCTAATGCT 74

RESULT 11
 LOCUS R52778/c
 DEFINITION y999c01.r1 Soares infant brain INTB Homo sapiens cDNA clone
 IMAGE:41742 5', mRNA sequence.
 ACCESSION R52778
 NID 9814680
 VERSION R52778.1 GI:814680
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 496)
 Hillier, L., Lennon, G., Becker, M., Ronaldo, M.F., Chapel, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, M.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,
 Marais, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Scheilenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
 Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
 and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478

TITLE
 JOURNAL
 MEDLINE
 COMMENT

CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 Insert Size: 1524
 High quality sequence stops: 340 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1524 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 340.
 Location/Qualifiers
 1..496
 /organism="Homo sapiens"
 /db_xref="GDB:414283"
 /db_xref="taxon:9606"
 /clone="IMAGE:41742"
 /clone_1lb="Soares infant brain INTB"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: lambda B; Site 1: Not
 I; Site 2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer [5';
 AACTGAGAGATTCGCGCCGAGAGATTTTCTTTTCTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (Pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the lambda B vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 149 a 117 c 114 g 112 t 4 others

ORIGIN

Query Match 79.0%; Score 15.8; DB 22; Length 496;
 Best Local Similarity 89.5%; Pred. No. 6.7e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTCTAATCTAATGCT 19
 DB 343 TGTCTAATCTAATGCT 325

RESULT 12
 LOCUS R59012/c
 DEFINITION y999c01.r1 Soares infant brain INTB Homo sapiens cDNA clone
 IMAGE:41331 5', mRNA sequence.
 ACCESSION R59012
 NID 9829707
 VERSION R59012.1 GI:829707
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 500)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P., and
 Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 On May 8, 1995 this sequence version replaced gi:800052.

TITLE
 JOURNAL
 COMMENT

CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 Insert Size: 1241
 High quality sequence stops: 396 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1241 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 396.
 Location/Qualifiers
 1..500
 /organism="Homo sapiens"
 /db_xref="GDB:413872"
 /db_xref="taxon:9606"
 /clone="IMAGE:41331"
 /clone_1lb="Soares infant brain INTB"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: lambda B; Site 1: Not
 I; Site 2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer [5';
 AACTGAGAGATTCGCGCCGAGAGATTTTCTTTTCTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (Pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the lambda B vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 137 a 145 c 127 g 86 t 5 others

ORIGIN

Query Match 79.0%; Score 15.8; DB 22; Length 500;
 Best Local Similarity 85.0%; Pred. No. 6.7e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTCTAATCTAATGCT 20
 DB 441 TGTCTAATCTAATGCT 422

RESULT 13
 LOCUS N50767
 N50767 214 bp mRNA EST 14-FEB-1996

DEFINITION VY90f02.r1 Soares multiple sclerosis 2NBMSP Homo sapiens cDNA
Clone IMAGE:280827 5', mRNA sequence.

ACCESSION N30767
NID g1191933
VERSION N50767.1 GI:1191933
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 214)
Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Ronfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The Mashu-Merck EST Project
Unpublished (1995)
On Apr 14, 1993 this sequence version replaced gi:785418.

TITLE JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: T7
High quality sequence stop: 180.

FEATURES
Source
Location/Qualifiers
1..214
/organism="Homo sapiens"
/db_xref="GDB:389203"
/db_xref="taxon:9606"
/map="10"
/clone="IMAGE:280827"
/clone_lib="Soares_multiple_sclerosis_2NBMSP"
/sex="male"
/tissue_type="multiple sclerosis lesions"
/dev_stage="Age 46"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker V-type; phagemid; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAAGTGGAGCGCGCCGATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot - 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH).

BASE COUNT 53 a 47 c 44 g 61 t 9 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 25; Length 214;
Best Local Similarity 89.5%; Pred. No. 8.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCCTGAATCTATGCT 19
||||| ||| |||
Db 176 TGCCTGAATCTATGCT 194

RESULT 14
AA020934 642 bp mRNA EST 30-JAN-1997
LOCUS AA120789
DEFINITION z64fi2.r1 Soares retina N2b4HR Homo sapiens cDNA clone
IMAGE:363791 5', mRNA sequence.

ACCESSION AA020934
NID g1484696
VERSION AA020934.1 GI:1484696
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 642)
Hallier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hulman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Marra, M., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Ronfing, T., Scheilenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
On May 8, 1995 this sequence version replaced gi:799442.

TITLE JOURNAL
MEDLINE
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1229 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 412.

FEATURES
Source
Location/Qualifiers
1..642
/organism="Homo sapiens"
/db_xref="GDB:128049"
/db_xref="taxon:9606"
/clone="IMAGE:363791"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). The retina was obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 185 a 119 c 148 g 183 t 7 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 27; Length 642;
Best Local Similarity 89.5%; Pred. No. 6.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCCTGAATCTATGCT 19
||||| ||| ||| |||
Db 175 TGTGTGAATCTATGCT 193

RESULT 15
AA120789 305 bp mRNA EST 19-NOV-1996
LOCUS AA120789
DEFINITION zK90d11.r1 Soares pregnant_uterus_NBHPU Homo sapiens cDNA clone

IMAGE:480101.5', mRNA sequence.

ACCESSION A1120789

NID 91677984

VERSION A1120789.1 GI:1677984

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 305)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holmen, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevasakis, E., Waterston, R., Williamson, A., Woldmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

On May 8, 1995 this sequence version replaced gi:800951.

COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 255.

Location/Qualifiers

1..305

/organism="Homo sapiens"

/db_xref="GDB:380458"

/db_xref="taxon:9606"

/clone_lib="IMAGE:490101"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;

Site_2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5'

AACGGAAGATTGCGGCGCCCTTTTCTTTTCTTTT 3'].

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization. Library

constructed by M. Fatima Bonaldo."

BASE COUNT 78 a 60 c 62 g 96 t 9 others

ORIGIN

Query Match 79.0%; Score 15.8; DB 28; Length 305;

Best Local Similarity 89.5%; Pred. No. 7.8e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCCTGTAATCTAATGCT 19

|||||

Db 254 TCCTGTAATCTAATGCT 272

Search completed: September 13, 1999, 15:45:56

Job time: 4044 sec

CDS

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGAR-1"
1..3207
/EC_number="2.7.1.67"
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="AA030698.1"
/db_xref="PID:9163520"
/db_xref="GI:163520"
/translation="MPPRSSGELMGHILMPRIIVLVECLIPNGMIVTLECLREATLIT
IKHELFKEARKYPLHQLQDESSYIFVSVTOEAREBEDEFETRCLDLRLOPFLKVI
EYGNREKILNREIGFALGMPCVCEPDNDPEODFRNIIANCKEAVDLNLSPH
SRAMVYPPNVESSPELPRKHTYINKDKGCIIVYIIVYSPNDKRTYLLKINHCVPE
OIAEAIKRTKRSMLLSSEQLKCVLEQKTLVCGDEFLKRTYLLKINHCVPE
IMLGMPNIMLMAKESLSQLEPMDCFTMPSYSRSTAIYPANGETSKSLAVINSAL
RIKICATYVANNIRIDIKIYVGTIGYHGEGLCNVTOQVPCSNPRNENLNDIY
IPDLPRARLCLSTICSVKGRKAEKHEBPLAMGNNINLEDTITLVSGKMLNLPVPH
GLEDLNPFGVTSNPKETPCLEEDPMFSSVVPFPMVSVEEHNANVSREAGFSY
SHAGLSNRLADNEIRENDKEQLAICTRDLSETTEBOKDFLMSHRHICYTIPILP
KILSVKNNSRDEVAQMYCLVNDMPPIKPEQAMELLDNYPDPWVGFAVRCLEYLT
DOKLSQYLIQVQVAKYEOYLDNLLVRLKALTNORIGHEFWHLKSEKNTVSO
RFGILLESYCRACGMYLKHLNROYEAMEKILNLDILQEKDETOYKQMFLEOMR
RPEDNALOGFLSPNPAHOLGNLEECRISSAKRPLMWNENDPDIASELTPONE
IIFKNGDDLRODMTLQIIRIMENTWONGDILRMLPGCLSTICGCVLIEVNSHT
IMQIOCKGKLGALQFNSHTLHQLKDKNKEIYDAIDLFTRSCAGICVAFILIGIG
DRHNSNINWKDQGLFHDGHEFLDHKKRKGKRYRERPEVLTODFLVISKAGQECT
KTREPEROEKCYAYLAIRHNALEINLFEMSLGSGMPELOSFDDIAYIRKTLALDK
TEOEALVEFMQNDAHNGWTTKMDTIFHTRKHALN"
BASE COUNT      1028 a      581 c      680 g      918 t
ORIGIN

```

```

Query Match      100.0%; Score 20; DB 3; Length 3207;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Cy 1 GTAATTCATGAACAATGA 20
    ||||||||||||||||
Db 3111 GTAATTCATGAACAATGA 3130

```

```

RESULT 2
LOCUS      A37232      3498 bp      DNA      PAT      05-MAR-1997
DEFINITION Sequence 1 from Patent WO9403609.
ACCESSION A37232
NID        92294345
VERSION    A37232.1 GI:2294345
KEYWORDS
SOURCE     unidentified.
           unidentified.
           unclassified.
REFERENCE 1 (bases 1 to 3498)
AUTHORS   Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
TITLE     EDUCAROTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR         OR PROTEIN KINASE AND ASSAYS USING THEM
PATENT:   WO 9403609-A 1 17-FEB-1994;
          IMP CANCER RES TECH (GB)
          Other publication JP 8503124T 960409.
FEATURES
SOURCE
1..3498
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
1..3207
/note="unnamed protein product"
/codon_start=1
/protein_id="CA002284.1"
/db_xref="PID:e304808"
/db_xref="PID:92294346"
/db_xref="GI:2294346"
/translation="MPPRSSGELMGHILMPRIIVLVECLIPNGMIVTLECLREATLIT
IKHELFKEARKYPLHQLQDESSYIFVSVTOEAREBEDEFETRCLDLRLOPFLKVI

```

```

EYGNREKILNREIGFALGMPCVCEPDNDPEODFRNIIANCKEAVDLNLSPH
SRAMVYPPNVESSPELPRKHTYINKDKGCIIVYIIVYSPNDKRTYLLKINHCVPE
OIAEAIKRTKRSMLLSSEQLKCVLEQKTLVCGDEFLKRTYLLKINHCVPE
IMLGMPNIMLMAKESLSQLEPMDCFTMPSYSRSTAIYPANGETSKSLAVINSAL
RIKICATYVANNIRIDIKIYVGTIGYHGEGLCNVTOQVPCSNPRNENLNDIY
IPDLPRARLCLSTICSVKGRKAEKHEBPLAMGNNINLEDTITLVSGKMLNLPVPH
GLEDLNPFGVTSNPKETPCLEEDPMFSSVVPFPMVSVEEHNANVSREAGFSY
SHAGLSNRLADNEIRENDKEQLAICTRDLSETTEBOKDFLMSHRHICYTIPILP
KILSVKNNSRDEVAQMYCLVNDMPPIKPEQAMELLDNYPDPWVGFAVRCLEYLT
DOKLSQYLIQVQVAKYEOYLDNLLVRLKALTNORIGHEFWHLKSEKNTVSO
RFGILLESYCRACGMYLKHLNROYEAMEKILNLDILQEKDETOYKQMFLEOMR
RPEDNALOGFLSPNPAHOLGNLEECRISSAKRPLMWNENDPDIASELTPONE
IIFKNGDDLRODMTLQIIRIMENTWONGDILRMLPGCLSTICGCVLIEVNSHT
IMQIOCKGKLGALQFNSHTLHQLKDKNKEIYDAIDLFTRSCAGICVAFILIGIG
DRHNSNINWKDQGLFHDGHEFLDHKKRKGKRYRERPEVLTODFLVISKAGQECT
KTREPEROEKCYAYLAIRHNALEINLFEMSLGSGMPELOSFDDIAYIRKTLALDK
TEOEALVEFMQNDAHNGWTTKMDTIFHTRKHALN"

```

BASE COUNT

```

1147 a      626 c      722 g      1003 t
ORIGIN

```

```

Query Match      100.0%; Score 20; DB 5; Length 3498;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Cy 1 GTAATTCATGAACAATGA 20
    ||||||||||||||||
Db 3111 GTAATTCATGAACAATGA 3130

```

```

RESULT 3
LOCUS      HSPH13K      3424 bp      mRNA      PRI      24-AUG-1995
DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
ACCESSION Z29090
NID        9472990
VERSION    Z29090.1 GI:472990
KEYWORDS   phosphatidylinositol 3-kinase.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
           Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1068)
AUTHORS   Volinia,S., Hilsen,I., Ormondtroyd,E., Nizetic,D., Antonacci,R.,
           Rocchi,M. and Waterfield,M.D.
TITLE     Molecular cloning, cDNA sequence, and chromosomal localization of
           the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene
JOURNAL   Genomics 24 (3), 472-477 (1994)
MEDLINE   95229146
REFERENCE 2 (bases 1 to 3424)
AUTHORS   Volinia,S.
TITLE     Direct Submission
JOURNAL   Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig
           Institute for Cancer, Research, 91 Riding House Street, London, W1P
           8BT, UK

```

FEATURES

SOURCE

CDS

```

Location/Qualifiers
1..3424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="KG1a"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CA02333.1"
/db_xref="PID:9472991"
/db_xref="GI:472991"
/translation="MPPRSSGELMGHILMPRIIVLVECLIPNGMIVTLECLREATLIT
IKHELFKEARKYPLHQLQDESSYIFVSVTOEAREBEDEFETRCLDLRLOPFLKVI
EYGNREKILNREIGFALGMPCVCEPDNDPEODFRNIIANCKEAVDLNLSPH
SRAMVYPPNVESSPELPRKHTYINKDKGCIIVYIIVYSPNDKRTYLLKINHCVPE
OIAEAIKRTKRSMLLSSEQLKCVLEQKTLVCGDEFLKRTYLLKINHCVPE
IMLGMPNIMLMAKESLSQLEPMDCFTMPSYSRSTAIYPANGETSKSLAVINSAL
RIKICATYVANNIRIDIKIYVGTIGYHGEGLCNVTOQVPCSNPRNENLNDIY
IPDLPRARLCLSTICSVKGRKAEKHEBPLAMGNNINLEDTITLVSGKMLNLPVPH
GLEDLNPFGVTSNPKETPCLEEDPMFSSVVPFPMVSVEEHNANVSREAGFSY
SHAGLSNRLADNEIRENDKEQLAICTRDLSETTEBOKDFLMSHRHICYTIPILP
KILSVKNNSRDEVAQMYCLVNDMPPIKPEQAMELLDNYPDPWVGFAVRCLEYLT
DOKLSQYLIQVQVAKYEOYLDNLLVRLKALTNORIGHEFWHLKSEKNTVSO
RFGILLESYCRACGMYLKHLNROYEAMEKILNLDILQEKDETOYKQMFLEOMR
RPEDNALOGFLSPNPAHOLGNLEECRISSAKRPLMWNENDPDIASELTPONE
IIFKNGDDLRODMTLQIIRIMENTWONGDILRMLPGCLSTICGCVLIEVNSHT
IMQIOCKGKLGALQFNSHTLHQLKDKNKEIYDAIDLFTRSCAGICVAFILIGIG
DRHNSNINWKDQGLFHDGHEFLDHKKRKGKRYRERPEVLTODFLVISKAGQECT
KTREPEROEKCYAYLAIRHNALEINLFEMSLGSGMPELOSFDDIAYIRKTLALDK
TEOEALVEFMQNDAHNGWTTKMDTIFHTRKHALN"

```

INLGRMNLKMAKESLYSOLPMDCTMPSVRSISATATPVNGETSTKSLVATRAL
 RITLCAATYNNLRNDIKYIVRGITGEGEPICDNYRTQVPSNRRNEMLNDIT
 IPDLPRARICLVSTCSYKGRKGAKEEHCPLWGNINLPDITDILVSGKMLNMPVH
 GLEDDLNPIDIGVTSNPKETPCLEEDMFSSVYKPPDMVIEHAWMSYREAGFSY
 SHAGLSNRLARDNLRNDKDEQLAISTROPLESTITQEOFLSHRHCTTIEIIP
 KLLSVNSRDEVAQMYCLVKMPPIKPPQAMELLDCNPDPMVREGAVACLEKYL
 DDKLSOYLIOVLKYEOYLDNLVRELKALTNORIGHEFFHAKSEMHNTVSO
 RRGLLSEYCRACQMYLKLHNRQYEAHEKILNLDLIKOEKEDTOKYOKFLEVQNR
 RPDMDALQGLSLPNPAHOLGNRLKECRIMSSAKRPLMNEMPDIMSELLEPQNE
 IIFKNGDLDQMDLTQIIRIMENIMONOGDLMLPYGLSLGDCVGLLEVNSHT
 IMOJOCKGGLKALQENSHTLHOMLKDNKNGEITDAIDLEFRSCAGCYATFLLGIG
 DRHNSNTMVDGQLFHDGFHLDHKKRKEGYRERVPYLLQDFLIVISKGQECT
 KTRFERFOEMCYKAYLAIRQANLFTNLSFMGSGMPPELQSFDDIAYIRKTLALDK
 TEQALTEYFKQMDAHHGWTTKMDIFHTIKQHALN"

BASE COUNT 1134 a 618 c 709 g 963 t

ORIGIN

Query Match 100.0%; Score 20; DB 10; Length 3424;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTATTTCATGAACAATGA 20
 ||||||||||||||||
 Db 3123 GTATTTCATGAACAATGA 3142

RESULT 4
 HS079143 3207 bp mRNA PRI 02-JAN-1997
 LOCUS Human phosphoinositide 3'-hydroxykinase p110-alpha subunit mRNA,
 DEFINITION complete cds.
 ACCESSION U79143
 MID G1763625
 VERSION U79143.1 GI:1763625
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 3207)
 AUTHORS Stirdivant,S.M., Ahern,J., Conroy,R.R., Barnett,S.F., Ledder,L.M.,
 Oliff,A. and Heimbrock,D.C.
 TITLE Catalytic Activity of the p110-alpha Subunit of Human
 Phosphoinositide 3'-Hydroxykinase is Required for Signal
 Transduction
 JOURNAL Bioorg. Med. Chem. (1996) In press
 REFERENCE 2 (bases 1 to 3207)
 AUTHORS Stirdivant,S.M., Ahern,J., Conroy,R.R., Barnett,S.F., Ledder,L.M.,
 Oliff,A. and Heimbrock,D.C.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs,
 Summeytown Pike, West Point, PA 19486, USA
 FEATURES
 source location/Qualifiers
 1..3207
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..3207
 /function="phospholipid kinase"
 /codon_start=1
 /product="phosphoinositide 3'-hydroxykinase p110-alpha
 subunit"
 /protein_id="AAB39753.1"
 /db_xref="PID:G1763625"
 /db_xref="GI:1763625"
 /translation="MPPRSGEIMGTIHPRIPLIVCEILPNGMIVTLECEATLLT
 IKHELEKARKYPLHOLLQDESSITLYSVQEARREPERFDTRLCDRLQPLKVI
 EPVGNREKILNREIGFALIGPVECFVADKPEVODFRNRLNVCKRAVDRLDINSPE
 SRAYVVPNVESSPELPKHLIYNRLDKGOIIVVIVVSPNNDKQKYLKINHCVSE
 OVIAEARKTRSMILSSEQLKCLVLEYGQKYLKVGCDYEPLEKYLISQYKTRSC
 IMGRMNLKMAKESLYSOLPMDCTMPSVRSISATATPVNGETSTKSLVATRAL
 RITLCAATYNNLRNDIKYIVRGITGEGEPICDNYRTQVPSNRRNEMLNDIT
 IPDLPRARICLVSTCSYKGRKGAKEEHCPLWGNINLPDITDILVSGKMLNMPVH
 GLEDDLNPIDIGVTSNPKETPCLEEDMFSSVYKPPDMVIEHAWMSYREAGFSY
 SHAGLSNRLARDNLRNDKDEQLAISTROPLESTITQEOFLSHRHCTTIEIIP
 KLLSVNSRDEVAQMYCLVKMPPIKPPQAMELLDCNPDPMVREGAVACLEKYL
 DDKLSOYLIOVLKYEOYLDNLVRELKALTNORIGHEFFHAKSEMHNTVSO
 RRGLLSEYCRACQMYLKLHNRQYEAHEKILNLDLIKOEKEDTOKYOKFLEVQNR
 RPDMDALQGLSLPNPAHOLGNRLKECRIMSSAKRPLMNEMPDIMSELLEPQNE
 IIFKNGDLDQMDLTQIIRIMENIMONOGDLMLPYGLSLGDCVGLLEVNSHT
 IMOJOCKGGLKALQENSHTLHOMLKDNKNGEITDAIDLEFRSCAGCYATFLLGIG
 DRHNSNTMVDGQLFHDGFHLDHKKRKEGYRERVPYLLQDFLIVISKGQECT
 KTRFERFOEMCYKAYLAIRQANLFTNLSFMGSGMPPELQSFDDIAYIRKTLALDK
 TEQALTEYFKQMDAHHGWTTKMDIFHTIKQHALN"

GLEDDLNPIDIGVTSNPKETPCLEEDMFSSVYKPPDMVIEHAWMSYREAGFSY
 SHAGLSNRLARDNLRNDKDEQLAISTROPLESTITQEOFLSHRHCTTIEIIP
 KLLSVNSRDEVAQMYCLVKMPPIKPPQAMELLDCNPDPMVREGAVACLEKYL
 DDKLSOYLIOVLKYEOYLDNLVRELKALTNORIGHEFFHAKSEMHNTVSO
 RRGLLSEYCRACQMYLKLHNRQYEAHEKILNLDLIKOEKEDTOKYOKFLEVQNR
 RPDMDALQGLSLPNPAHOLGNRLKECRIMSSAKRPLMNEMPDIMSELLEPQNE
 IIFKNGDLDQMDLTQIIRIMENIMONOGDLMLPYGLSLGDCVGLLEVNSHT
 IMOJOCKGGLKALQENSHTLHOMLKDNKNGEITDAIDLEFRSCAGCYATFLLGIG
 DRHNSNTMVDGQLFHDGFHLDHKKRKEGYRERVPYLLQDFLIVISKGQECT
 KTRFERFOEMCYKAYLAIRQANLFTNLSFMGSGMPPELQSFDDIAYIRKTLALDK
 TEQALTEYFKQMDAHHGWTTKMDIFHTIKQHALN"

BASE COUNT 1043 a 586 c 670 g 908 t

ORIGIN

Query Match 100.0%; Score 20; DB 10; Length 3207;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTATTTCATGAACAATGA 20
 ||||||||||||||||
 Db 3111 GTATTTCATGAACAATGA 3130

RESULT 5
 AF029224 12816 bp DNA BCT 23-APR-1999
 LOCUS Staphylococcus carnosus nlr and nar operons, complete sequences.
 DEFINITION AF029224 AF029225
 ACCESSION G4596720
 MID AF029224.2 GI:4596720
 VERSION AF029224.2 GI:4596720
 KEYWORDS Staphylococcus carnosus.
 SOURCE Staphylococcus carnosus
 ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Staphylococcus.
 REFERENCE 1 (bases 6052 to 12816)
 AUTHORS Pantel,I., Lindgren,P.E., Neubauer,H. and Goetz,F.
 TITLE Identification and characterization of the Staphylococcus carnosus
 nitrate reductase operon
 JOURNAL Mol. Gen. Genet. 259 (1), 105-114 (1998)
 REFERENCE 2 (bases 1 to 6064)
 AUTHORS Neubauer,H., Pantel,I. and Goetz,F.
 TITLE Molecular characterization of the nitrite-reducing system of
 staphylococcus carnosus
 JOURNAL J. Bacteriol. 181 (5), 1481-1488 (1999)
 MEDLINE 99175446
 JOURNAL 3 (bases 1 to 12816)
 REFERENCE Neubauer,H., Pantel,I. and Goetz,F.
 AUTHORS Submitted (26-SEP-1997) Mikrobielle Genetik, Universitaet
 Tübingen, Waldhauser Str. 70/8, Tübingen 72076, Germany
 JOURNAL On Apr 21, 1999 this sequence version replaced gi:3929521
 COMMENT gi:4433635.
 FEATURES
 source location/Qualifiers
 1..12816
 /organism="Staphylococcus carnosus"
 /strain="TM300"
 /db_xref="taxon:1281"
 226..1056
 /gene="nlrC"
 226..6032
 /gene="nlr operon"
 226..1056
 /gene="nlrC"
 /function="putative nitrite transporter"
 /codon_start=1
 /transl_table=11
 /product="nlrC"
 /protein_id="AAD0822.1"
 /db_xref="PID:G4433636"
 /db_xref="GI:4433636"

```
/translation="MWOMKHTVEDTYASRETMNSVSSOMKSYMLSTKPYFLKA
MMSGFLSVAVFEMAIKIQHNGINDGTINLGAIAFSAIULIYITSELTSMFMY
MTAGATYKSVPIGKIWTLPYTCIGNIVGGFVLGULKTISVTPMEMYALATKYDK
TTLGTWGDILVAIFANFFIKIDIVYSIQFKGLSKAFIACGVVVFVFMGEHYVN
ANLFSGMEFYNDALSMIDVLKNIMAFVAGNYIGGIFVGLLYAFNCRNHYIDKOK
"
gene      1184..1906
CDS       /gene="nirR"
          1184..1906
          /function="Involved in nitrite reduction"
          /codon_start=1
          /transl_table=11
          /product="NirR"
          /protein_id="AAD20823.1"
          /db_xref="PID:94433637"
          /db_xref="GI:4433637"
          /translation="MSQTIIVHMKRGKRONOLMELEHVKOYNGEEDVAFISDE
          LSLPVIRKYLAEHGEHPTIPLLLFSGHRYEDIPTLMKEMQLMPEGSIHQIROP
          L CYHPALTNKIDAKKTONNSNAIYVIGSGSPFLTPEDDELMLKKECTECYV
          TMETYGDOFEKLEPKMDNYHLLITPEFYDGLVKNIKKINDINGSDMTAHA
          LNFEPIDEMLAARLNEVKEFS"
          1907..2362
          /gene="sira"
          1907..2362
          /note="similar to N-terminus of Escherichia coli CysG."
          /codon_start=1
          /transl_table=11
          /product="Sira"
          /protein_id="AAD20824.1"
          /db_xref="PID:94433638"
          /db_xref="GI:4433638"
          /translation="WPIQLNKNKRAVALLGGGRIGFRKQOLANADGCVTVISKAP
          WPEFHSYPIGKIITKDYKEDIADADAVIATDSOEINIKODTTPQOVLNHTGD
          KSDSFMNIEFDLAIYSVNSGDKYKQVSKAIKRYIQEYGR"
          2363..4771
          /gene="nirB"
          2363..4771
          /note="similar to Escherichia coli NADH-dependent nitrite
          reductase large subunit NirB."
          /codon_start=1
          /transl_table=11
          /product="NirB"
          /protein_id="AAD20825.1"
          /db_xref="PID:94433639"
          /db_xref="GI:4433639"
          /translation="MSKORLVIGNGMAGVRIETIIDRPDKEDITIIIGKPPNRY
          RIMLSNIIQKMTVEETIMNPSWYENGIALITGEKATRVNREKHIVETTERGRI
          DICIFATGSKAFVLPFGHELPSVIGMRTIEDTEKLEINODKKEATYVIGGGLLEPC
          AGLDQGMNIVVLAELMLMOIDROAGDMLRDLLEQKRFELGANTEIIGNOD
          VEALIRLADGRDLPADLVVAVGIRKPTDEAIDSGIDVREGIYVDFMRTSPDEIYAVG
          ECAQHNKCYVGLAVLYEQGVLAHLTGREGYOGSTFSLKYSGLDLSAQIO
          EDDEVEGEIYNSHDHYKRIKLEERYVAVLYGDTDDGTFEFNMNMRKEPIEDYTL
          VSLQKGEENAVTADMDSETTIGCGNGVSGKIVSAIVEDGLTSAVETKATAGN
          SCGRKPOIGELLEHTLGGAFYASKPAGICDTELTROIVQITAKGIKTSKEVRY
          LDMNKGKCPKPCRIINITYLNNVYPHEHEDEKASPPANDERIANIONGTSVIOMR
          GGVTPDQILRIGEVAKRYDVLVTVGSORIGLISPEIDLPKWKDLGMSASAVA
          KTRISVSKSCVCKEFCFGTQYTRIGIRLEETFEYIDTPHKFMGVSQPCRSQVSG
          KDFGVIGENGFOIYIGNGGTEVTKGOFILVEVEDYICGAYMOYRETGYAE
          RTAPWLERMGFENVEYVLDPEKQALDRIDAKARAVAEPEMSEIVDQDKIKFY
          EKV"
          4772..5088
          /gene="nirD"
          4772..5088
          /note="similar to Escherichia coli NADH-dependent nitrite
          reductase small subunit NirD."
          /codon_start=1
          /transl_table=11
          /product="NirD"
          /protein_id="AAD20826.1"

gene      5079..6032
CDS       /gene="sira"
          5079..6032
          /note="similar to C-terminus of Escherichia coli CysG."
          /codon_start=1
          /transl_table=11
          /product="Sira"
          /protein_id="AAC82542.1"
          /db_xref="PID:93929522"
          /db_xref="GI:3929522"
          /translation="MAKGMNFRKPTKENGMSVLTDKSNEMKMYREMSHDKYVR
          TTHGNCVGSCKAVFYKNGYITNENQDTPVPSGPDMPPEFPGCRGASFSYTES
          PLRVPIYPMRKLMEIWTALIEHEGNNIAWASIVDPEKAKIYNKARGGHNRSYS
          WKDVLQILAQGLLITIKKDGPRLAGFTPIPAMSLVASCARFINLNGKBLSEFYW
          YADLPASPOJMGEDQTDVPESSDMYNNAYIIMGSNPLRTPPAHFTEVRYGCTYV
          VSAVDYENKVFADNMLAPRNGDAVAOAMTHVILOKYVDESEFVYAKOYTD
          MPILMPEHHEGKAGRPILASDLGSEENEMKPYIVDSEGITIIPATMGORRE
          EGKONKLTENEDSKYDPRALSVABGEDDIDIMKFPFPDANKGIFERTPIASRYTTLA
          DGRKVAATVLDLTSQIGIRFNHELEAKFDEATYTPAMOEAAVAGVQDVIYTOV
          ATERAQNADIDGSRMILMGAINHMSDSYIYAVLVNLWMLCCOGNGGMAHYVG
          QEKCRPIEGMSTIAFAKDMQGPRLQNSWTFYFTTQMKYEEGCVKALPLAGELK
          HKHPADYVNTAARGLMPLSPYQFOTNSLRFEGDEAKEGEFNEEVLKAAVSVSRET
          KEAVDPRLRTNHRKSLFIWRSNLSSAGOEYFMKMTGSSGILLAEPEDEKPRE
          MIRWDVQKIDILVALDFRNTSTPLKADYILPRAIYTERKHDLSITDMHPPVHPNPA
          IDPLMESRSDMDIRKSQVYSEMAKYMPTVYDVTSPLAHDSKQELIATPYKVD
          WTKEVESIPKRTYMGFAVVDRTLLYDKVYTAGPLLETAKIAGHGVSPVFKOYEE
          AKDIDVGSWSDTITIDGRPIDTARADAVYINISANGELSDASDYLEKTCOMPLT
          DISDSRSEKRTFNNTISOPREVIPTAVFSGSNKNRYSPFTNINELVAFRTLGR
          OSYVILHIFQOFSGLPYVYKPTLPVAVFTGKQVYGGQONLVLRLTLPHGKNHS
          TYQONQHLITFRGCPPIYAKLEDAEHHDHNDMLDYNNRNGVYITARAATYSHRPG
          TMTFTHADKRIQTGSEFTGLRGSHAPRIHLRTPQLVGTQIYSTSNITGPIS
          NQROVYVAVRKKEVMDLED"
          9989..11566
          /gene="narH"
          9989..11566
          /note="similar to Escherichia coli nitrate reductases NRA
          and NR2 beta subunits."
          /codon_start=1
          /transl_table=11
          /product="NarH"
          /protein_id="AAC82543.1"
          /db_xref="PID:93929523"
          /db_xref="GI:3929523"
          /translation="MKIKAQVAVVNLNDCICGHTCSVTKNTNTNRPAGAYIWFNNV
          ETKPGVGYPKRMEDEHYKGGMTLNKGIKDLKSGWSKIAFGKITYNPDMPIYIKY
          YEPWTYVNTNHTTAKASKHSPVATAHSAVISGDKIDLMGWPMWEDDLAGHVTGDEPN
```


/note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB42162.1"
 /db_xref="PID:61429602"
 /db_xref="PID:64756154"
 /db_xref="GI:4756154"
 /translation="MNOFLWVTPYLCVAIFITGHANRYRDOFSWAKSSEETIKK
 RIKWGSILFHLGIIPIVAAAGHVGVLVPAALMEQGVNHHIYHIGAVYISLFGFMYLI
 GMILYRVSINKIRLSADNVNFFLVIIICGLSTVTNNAISPEFVROFIS
 IMFRHLFTNPNASLMTNVPMSFKLHIFIGFSIFALMPFRLVHVMVSPLSYANRSYI
 IYRKNKRR"

BASE COUNT 3550 a 1771 c 2304 g 3045 t

ORIGIN

Query Match 87.0% Score 17.4; DB 5; Length 10670;
 Best Local Similarity 94.7% Pred. No. 2.5e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATTTCATGAACAATG 19
 |||||
 Db 6284 GATTTCATGAACAATG 6302

RESULT 7
 ABO17071 81370 bp DNA PLN 06-FEB-1999
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MSJ11,
 DEFINITION complete sequence.
 ACCESSION ABO17071
 NID 93510347
 VERSION ABO17071.1 GI:3510347
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 REFERENCE 1 (sites)
 AUTHORS Nakamura, Y.
 TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. I
 JOURNAL Unpublished (1998)
 REFERENCE 2 (bases 1 to 81370)
 AUTHORS Nakamura, Y.
 TITLE Direct Submission
 JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases.
 Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
 Gene Structure 2, 1532-3, Yana, Kasarazu, Chiba 292, Japan
 (E-mail: ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
 Fax:+81-438-52-3934)
 FEATURES
 source 1. 81370
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="MSJ11"
 /clone_lib="Mitsui P1"

BASE COUNT 25619 a 14925 c 14478 g 26348 t

ORIGIN

Query Match 87.0% Score 17.4; DB 7; Length 81370;
 Best Local Similarity 94.7% Pred. No. 2.1e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TATTTCATGAACAATG 20
 |||||
 Db 14778 TATTTCATGAACAATG 14796

RESULT 8
 SCPP8GNA 8151 bp DNA PLN 17-FEB-1997
 LOCUS S.cerevisiae PRP8 gene, complete CDS.
 DEFINITION 224732
 ACCESSION 9395156
 NID 224732.1 GI:395156
 VERSION PRP8 gene; splicing factor.
 KEYWORDS baker's yeast.
 SOURCE Saccharomyces cerevisiae
 ORGANISM Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 8151)
 AUTHORS Hodges, P.E., Jackson, S.P., Brown, J.D. and Beggs, J.D.
 TITLE Extraordinary sequence conservation of the PRP8 splicing factor
 JOURNAL Yeast 11 (4), 337-342 (1995)
 MEDLINE 95304847
 REFERENCE 2 (bases 1 to 8151)
 AUTHORS Jackson, S.P., Lossky, M. and Beggs, J.D.
 TITLE Cloning of the RNAB gene of Saccharomyces cerevisiae, detection of
 the RNAB protein, and demonstration that it is essential for
 nuclear pre-mRNA splicing
 JOURNAL Mol. Cell. Biol. 8 (3), 1067-1075 (1988)
 MEDLINE 88216580
 REFERENCE 3 (bases 1 to 8151)
 AUTHORS Lossky, M., Anderson, G.J., Jackson, S.P. and Beggs, J.
 TITLE Identification of a yeast snRNP protein and detection of
 snRNP-snRNP interactions
 JOURNAL Cell 51 (6), 1019-1026 (1987)
 MEDLINE 88080460
 REFERENCE 4 (bases 1 to 8151)
 AUTHORS Hodges, P.E.
 TITLE Direct Submission
 JOURNAL Submitted (21-JUL-1993) Hodges P. E., University of Edinburgh,
 Institute of Cell and Molecular Biology, Darwin Building, King's
 Buildings, Mayfield Road, Edinburgh, Midlothian, United Kingdom,
 EH9 3JR

FEATURES
 source 1. 8151
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /chromosome="VIII, right arm"
 /clone="PR800"
 /clone_lib="Sau3A I genomic DNA in YEp24, D.Botstein
 (MIT)"
 1..1318
 /note="overlapping sequence with CDC23 gene (GenBank
 D00610)"
 431..7901
 /gene="PRP8"
 /note="5' and 3' ends mapped by nuclease S1 protection.
 contains no introns"
 /citation=[1]
 /citation=[2]
 /citation=[3]
 /db_xref="SGD:S0001208"
 /function="splicing factor, U5 snRNP protein"
 /evidence="experimental"
 431..7901
 /gene="PRP8"
 /db_xref="SGD:S0001208"
 531..7772
 /gene="PRP8"
 /function="pre-mRNA splicing factor, U5 snRNP protein"
 /note="presumed C.elegans homologue encoded by C50C3.6
 (GenBank U14433)"
 /citation=[1]
 /citation=[2]
 /citation=[3]

```

/codon_start=1
/db_xref="SGD:S0001208"
/evidence-experimental
/product="Prp8"
/protein_id="CAA80854.1"
/db_xref="pid:g395157"
/db_xref="gi:395157"
/db_xref="SWISS-PROT:P33334"
/translation="MSGLPPEPGEEDSDLAPPPPPPEYEIEELDNPVSSV
EDTLPPEPPPSNEINAEIVDFLLPPEPPGIDLETNAEKVELAGKRLDIG
KDTETPKRSKRKAKKTKKAKSNLYTPKAEPPPHLRRIITHSDMSKMYTKKA
FLGALKYPLHALIKLENNPHEPQAKKVKYLYHSGATTFNENPPEVPEPTQWS
ATWLMREKXDRTEKRNRPPEDDDEPPSYENHTEPDELDPINPLSDQDEYV
KDWLDSRPLEEDSKVNGSTKMSFDPEKSNLRUSTPRLDEVDYDNTYLYDKK
SFEKGNALNNALPGCFKPEPLYPREEDYDNFNSIDRYFVRPIRSEKVAPELYN
SPRSVRIPWNNPSCIIONDEYDTPALFEPDINPIPHEDNNSSYANTKNG
DETLPEDAPLLAEELILPNTKDMSLYHSPFPNPKGKMAADYALAKWFLQ
HPDEYPAVKVYQKLLKNYLNELHPTLPNNHKLTKSLKTKKXPCQOTTIWE
AGLOCRGCHNMLLIHRKGYLHLDYNNLKPRTLTTERKSKLGNSEFHMR
LTKMKLVDTNVOFRGLGVDAFQIADGHIHLNHTIGLTYRKLVMQIRACKD
LKHILYFKNRLGPGCGFWQPMRWMLNRLGYIPLELYIGNLITRQEGSNE
IVKTKKORLDAVYDLELNSYMDILKMEPSLRKAKRTILOHLEAMRCWKANIP
MDVPMAPIKKIIERYIKSKADAWSAHYRERIKGAHVEKTMVKNKNGRLRM
IKNEBOROIQKNGPEITPEATTIFSVWMLSRSPSPPEPPLATYKNDTITVL
ALFDKDYVASKVRNASERELALIEAYNDPHDYLNRKIKYLLTQRYKRYDTTM
ENYONISPVISYDPLEKITDADQLMTADQRLFPWIKPDSSELPPLVYKMTQ
GINNLSEIMDVSRGQAVLLETTLEMAEKIDFTLNLRLVIDPNADYITAKNV
VINFRMSHVNGYGLRGLKFSFIFQYGLVIDLLGQERATDLAPANNPEMQ
FKSKEVERAHPIRLYTRYLRYLMEHEDEGEELTDEYLAENDPENNENIGNR
KCPWDSMRHLIRDOVNLGRAVEMEIOSRVPISLTKMENAFVYSKNNPILFSM
CGFEYRILPROMEEVYNDEGVNDYDERTKQRAKAYKISEEIKFDSRTGIL
MASGSTTTKAAKNISLISLTYFREIYATEPDLILVAGEIRIQNRVYGLNSK
MPTRPAPVFTPKELGIGMISASHILIPASDLSMSKOTDGIHFRAGMHEDEKL
IPITERYITWENELDSQVWAEYATKROEALIQONRRLAELESGMDRGPISLT
FQDRHTLAYDRGHRIRREPKOYSLERNSPWTHSGDKMKNANAYRTDIOALGG
ETITLEHTLFGTGFNSWEGLEWKAQSPDSMOKRTGLSOTPNRPFITM
SPTIRANVYGVYOLDLGTGLFKGKIPYLKIQIFRAHLMOKIHESIYDFDQI
LDGELDVLIQISSVTEYHPRKSTKNSADITMSVHEWVSAPSLHEINDSKG
LITNKMFDVOLRYGDYSDHISRYRAKFLDYTDYDNNVSPSPGVAGIDLAAMY
DAYGNFMGLKPLIONSMTIKNAPALVYLERIRKGIQIQQSVQEPFLSSYAE
LENNDIKLEVDINRYRYVHTFEGNATKINCCIETLNKTKHLEFKIHTSVWA
GOKRLSOLAKMTAEVSAVLSLPKEPROQIITYTRAMLDPLVHMLDPENITARP
TELRLPSAASIDKLDVYKATKATPOVLPRIYDMDRISSTYASRTITLALK
TNESAKMILSDPTTTTKSTHLPSPFDEQITTESQRIILITEYKGVINISAL
TQTEIKDILIQONIKAPSVKQKMALEAREKONDEAAAGASVMYTKYINAGEL
IYVVASADYESQTESKNNMRSAIANLILYRLNIVYASADFEEDNNVYLPNLL
KKFLEISDVKIOVAFIYGSMAKDPKVKETITVYVQLGVSGVOISINPIDIDLP
DTEGELGMHTORBELKPMASAVATHSKLPADKRDCDISIFSPGVSLSAVN
LTDGYOMGEENKIDIMNLSGFEPTSTHQLSDRTITGFIIPSNGVYITMGCT
APNODGDNFKIGLPLETYNEMHNPVHFLQSELAGDELEAQIDVDS"
/gene="prp8-1"
7570
mutation
/gene="prp8-1"
7570
/note="type srp8.3"
/citation=[1]
/citation=[2]
/replacement="a"
/phenotype="temperature-sensitive pre-mRNA splicing
defect"
7566..7859
polyA_site
/gene="prp8"
/note="major site, mapped by nuclease S1 protection"
/citation=[1]
/db_xref="SGD:S0001208"
/evidence-experimental
7863
polyA_site
/gene="prp8"
/note="minor site, mapped by nuclease S1 protection"
/citation=[1]
/db_xref="SGD:S0001208"
/evidence-experimental
7866
polyA_site
/codon_start=1

```

```

Query Match
Best Local Similarity 94.7% Pred No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0;
Gaps 0;

QY 2 TATTTCATGAACAATGA 20
Db 5677 TATTACATGAACAATGA 5695
|||||
RESULT 9
YSCDBF3A YSCDBF3A 7845 bp DNA PLN 22-MAY-1995
LOCUS YSCDBF3A 7845 bp DNA PLN 22-MAY-1995
DEFINITION YSCDBF3A Saccharomyces cerevisiae DBF3 gene, complete cds.
ACCESSION L29421
VERSION L29421.1 GI:460041
KEYWORDS Saccharomyces cerevisiae (library: genomic) DNA.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
REFERENCE 1 (bases 1 to 7845)
AUTHORS Dol, A. and Dol, K.
TITLE Cloning and nucleotide sequence of the CDC23 gene of Saccharomyces
cerevisiae
JOURNAL Gene 91 (1), 123-126 (1990)
MEDLINE 90382687
REFERENCE 2 (bases 1 to 7845)
AUTHORS Shea, J.E., Toyn, J.H. and Johnston, L.H.
TITLE The budding yeast US snRNP Prp8 is a highly conserved protein which
links RNA splicing with cell cycle progression
JOURNAL Nucleic Acids Res. 22 (25), 5555-5564 (1994)
MEDLINE 95140615
FEATURES
source Location/Qualifiers
1..7845
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/tissue_lib="genomic"
/map="chromosome VIII right arm"
225..7466
/gene="DBF3"
/db_xref="SGD:S0001208"
225..7466
/gene="DBF3"
/note="ORF"
/codon_start=1

```



```

/ab_xref="PID:g458904"
/ab_xref="GI:458904"
/translacion="MTIVVYVYIRKRGLEDNRROCTINCTPGPRQMTVGVFSERASL
THQGEFVIRKODKODKSDFYVSGSLIDALVYSLVADESISRVOMSKQOIF
SDRIYVPLTADSVYGAFAVLDKLPSTSPQNYPMDESLIGSDASNNKIAEYER
ITVQDVLVILGCGPGSGTSLPGEHTRYLNNETKRVAMCGSDPKPBDRLTFTIP
VLADAKALCFVABESSKQNMHELPDLKNDLPALVYKLPGEKTSFVVEBAFGVQ
TKTF"
gene
complement(2399, .6967)
/gene="DNA2"
complement(2399, .6967)
/gene="DNA2"
/note="YHR164C"
/codon_start=1
/evidence="not_experimental"
/product="dna2p: DNA replication helicase"
/protein_id="AAB68010.1"
/ab_xref="PID:g458906"
/ab_xref="GI:458906"
/translacion="MGTPOKNRKSASISVSPAKTKEKTIQNSKAILSKQTRRK
KYAFINNLNGKNTKYSNASVLSIVSOVRNTSRKIDINKAVSKSVKQIPNSOVP
KREMSNLSRHDFDQDGMEEVYIKYSPIDRMDSPKTSAAEYSDYDQVNPST
PIYPNRLKYTLSEFNIQVPADVNLQIENGNGOVRPKPAEISRESLNIDLDLDT
EGDLTKPTTKESDSSPIKAPNVRKAEVNAEVDKADSTGSDSDGLDIT
OKYVERKSSQITIGNTNKGSAQSGCNDTKRGELDEBENDQAKTGMAEY
ENEDSNCOQIRKNEKLEIYNSDESDSLILENTOVEPATIEQDIDRKEKWS
DDLRIADSTLSAVALRAKSGAPDGVYRIVLSRVLEPKITQTKILCIGKQV
SSVVRHPVYILEEVGVHIEGKNIENKRLSDKPKRTOALANDNLVLDNGF
SATSVSGVGLCRSLIOMOFODRGPRLVMTLGNVHLELDISIKYKSHNKISME
ITIOKDLSELYSFSITICNEEIOYKELMKHANITLYPNKPYSKNSGCTEIS
SGTRTOPITISVVIDIENIWSPIYOLKGFQDLYVANEKNAKHLVPLEVKGSR
SVSEYVGLIYTLNLNDRIEPILEFLLYPTFRKMKRPPSVLKHILNLSRMRM
NFRHOLOEVGOASREPLPLDSSDCSFCIESCNLMLKLEDGTPESSGVSEB
FELLTINSONLANEYKEFTKYNDLIKRESSITCYNKELELDGDSRESGCLSG
LVSEVYEHETEGAVIYCFSRNRNNSOSMLSOJAANDFVLSDEBHCFCOOR
VOPINPAKICISYKRLNLNRDLKREGVTTIOSVSELSOSSLATQNLVYTRIK
NDIQOSLSLAFNLSLFLPVSQVQIDVDSKLCRTKRSBGNELISLVDNA
PKRDANDDPVPIPKLSKDTLNLNQEAIDKVRADVALILGPGTKYVIAEII
KIIVSEGRKVLTSYHSADVNDILIKRNTNISIMRGMKRVHPDOKYVPAVSX
SYNDLSKINSTSVATTCGINDILFTLNKEDDYILDEASOISMPVGLPARYN
RFIYMGDHYOLPVLKNDARLIGLESLFKTEKHPESVASELTQYRMCGDVLVS
NFIYDQKALCGNVEYFQSLIEPMPALSRYSRESANSKOMLEDILEPRKYVFLW
DNCPDITQSEKDNITNHEAEELTQCVESHLGVPCEIDIGVATLRAOURLKLF
NKVYDDELEITADQFGQKDKCIISVNRNSQALNGALKLELRVNVAMTRAKSL
IIGVSKETISVEIKSFVNLERNMVYMCMDALYKRFPRDSNAIDBARNGCGR
TGAKPITSKSFSDKPIKEILOEYES"
complement(7496, .14737)
/gene="PRP8"
/ab_xref="SID:S0001208"
complement(7496, .14737)
/gene="PRP8"
/note="YHR165C: Involved in the control of S phase and
slicing"
/codon_start=1
/ab_xref="SID:S0001208"
/product="PRP8: RNA splicing factor"
/protein_id="AAB68011.1"
/ab_xref="PID:g551320"
/ab_xref="GI:551320"
/translacion="MSGLPPEPPGFEDSDLALPPEPPPGVIEBELDNPMVSSN
EDTFLPPEPPSPFEINAEIYDFTLPPPEPPGDELEKRAKVELGSKRLDQIG
KDFEYRKSRKRAKMTKAKRSLYTPKMPPEHRLKILNHSDAVSKYKQADLQ
FLGALKYLPALIKLENNHPMEQAKEVLYTSGLITFVNTPVATSEVYTAQWS
ATYIAMREKRDRTHEKRMFPEDDEDEPLSYOHLNENLEPDLPIILPDSODEVY
KWLUDSRPLEEDSKYNGSYKWSFDLDEMSTLYFRIPRIBEDVNDKYVYFEDK
SFEGRKALNNATIGGPKPEPLYPREEEDYNERISYFRIPRIBEDVNDKYVYFEDK
SRPSVRIPIWNNVSCIQNDSEYDIPALFEDSLNIPDPIDNNSSLVNSKENG
DFTLPEDFAPLAEELIIPNTKDALSLHSPPEPRKGVKRAQADVLAQKMPQ
HPDEEYKVKVYQKLEKLVNLNELPLTNNKRLKSLANTYKFOOTLIDWYE
AGLOLCROGNNMLILHRGLYVLEHDYNNGLPTLTTRKSKRLGSLFAME
LTKMKLIVDTHVOFRIGNDAYLOLAGIYIILNHITQDGIYKRYKWHQIFACD
LKHITIKPKNNKCKGPGCGFWQDPAWYVLAFLGTLPLERITIGNILITROFBSRNE
IVKTTIKORLDAYIDELRNSVMDLILEMPESIROKARTITQHLSEAMCWMKANIP
```

```

WDYGPAPPIKTIERYIKKAAWYSAHYNERIKRCAHVEKTVKKNLGLTLRL
IKNEQERQIQKNGEPITPEEATLFSVAVWELERSSEPIFPEPLYKNDKILVL
ALBDKDVYASKVRLNLSEREELALIEAANDRHDNLNKKILQIRYKPYDITM
ENYONISPYVSPLEKIDTAYADQKRLFPNKRPSOSEIPLIYKWTM
GINNLSHTMVSQGSVILETTLQMAEKIDTLLNRLRLITVDNINADYITAKNN
VINKRMSHNKRGILRGKFAFTIYQYVLYDILLGEBERTDLAGPANNRPMVO
FKSEYKHAHPIRLYTRIDRTYMLTFHEDESEELDTLAEIPNDPNSIGYNK
KMPKDSRMLIRQDVNLGRAVWEIQSRPSTLSIKMENAVSVYSKNNPMLFSM
CGFEVRLIPRORMEVYSDNGVMDVDEKORTAKAYLKEEIKKDSIRIL
MAGSTFTFVAAKMTLSLFTYFREAVLVAEPLDILVKGERTLONKRGKLSN
MTPRPVAVYTPRELGIGMISASHLILPASPILSKONDGTETHRGKMTHEDEK
IPTIFRITTWENEPDSQRYMAEYATKREALQONRLAFEBELSGMRCIRISL
FORDRHTLADRGHRLRREKQYSLSNSFEPMWNSHDKLNNLNAVRYDLOALGG
IETILBETTLFKGTGFSNMEGLFWEKASGEFSDNYLDRIDISSYAFSRLTLRLAK
TELRLEPFSAMSIDKSDVYMKATEPOMVFNIDYDMLDRISSYAFSRLTLRLAK
TNESAKMILSDPTITIKSYHLMPSTDEOWITIEQMDLILTEFGKRYNLSAL
TOPEIKDILIGONIKASVROMAELEARSKONDEEAAGASTVAKTYTNAOGE
IYVYASADYSQTFSSKNERKALANTLILYELKINYSADPVEBQNYVYIPKRLI
KRIEISDVRIQVAAFTLYGKSADHPKVKIKTVVLPQLGHVGSQISIPDIDQIP
DTGELGLGHIHQVOTELKMASEVATHEKLEADKRRDIDISFSTPGSVLSAIN
LTBEGVOMGENDIMNVLECEGEPTFSTAOQLLSRITGNFIISGNWNTFMGT
AFNOEGYNKCIPILEFYNMRPHFLQFSELAGDELEAQQIDVES"
complement(14959, .16839)
/gene="CDC23"
/ab_xref="SID:S0001209"
complement(14959, .16839)
/gene="CDC23"
/note="YHR166C"
/codon_start=1
/ab_xref="SID:S0001209"
/product="Cdc23p: cell cycle protein"
/protein_id="AAB68012.1"
/ab_xref="PID:g458908"
/ab_xref="GI:458908"
/translacion="MNDSDQKLIHDIRIOLRAATFSLRWKTYGSKAAAPALAGIA
EATDVDTSLASLEPRLNKGYPKQFELIPQNGFGISEYLYLGLSTFLPAKED
RCVFPLKDVNPIPLKFLKSLKSLSMWKSQESMENTLITGKFTDDEYRANKGDGSG
```

Query Match 87.0%; Score 17.4; DB 8; Length 41664;
Best Local Similarity 94.7%; Pred. No. 2, 2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TATTCATGAACAATGA 20
|||||
Db 9591 TATTACATGAACAATGA 9573

RESULT 11
HS1044017
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HS1044017 122325 bp DNA PRI 09-OCT-1998
Human DNA sequence from clone 1044017 on chromosome Xp11.3-11.4
Contains GSS and STRs, complete sequence.
AL023875
93449106
AL023875.1 GI:3449106
HUG.
human.
human sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

REFERENCE
1 (bases 1 to 122325)
Griffiths, D.

JOURNAL
Submitted (09-OCT-1998) E-mail enquiries: hunquerry@anger.ac.uk
Clone requests: clonerequest@anger.ac.uk
On Aug 21, 1998 this sequence version replaced gi:3488011.
During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

IMPORTANT: This sequence is not the entire insert of clone 1044017. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 50A13 (Z92545) is at 122226 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/CHX>
1044017 is from the library RpiC5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR.pCYPAC2>.

FEATURES

```
source
1..122325
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="1044017"
/map="p11.3-11.4"
/clone1lb="RpiC5"
382..2592
repeat_region
  /note="LI repeat: matches 2060. .4298 of consensus"
  2788..2815
  /note="14 copies 2 mer ta 89% conserved"
  complement(2894..3421)
  /note="LI repeat: matches 5191. .4664 of consensus"
  3426..3539
  /note="LI repeat: matches 5214. .5330 of consensus"
  3555..3890
  /note="LI repeat: matches 1. .348 of consensus"
  4119..4351
  /note="LI repeat: matches 381. .586 of consensus"
  4430..8034
  /note="LI repeat: matches 589. .4151 of consensus"
  8030..10187
  /note="LI repeat: matches 3232. .5390 of consensus"
  10039..10924
  /note="LI repeat: matches 1. .908 of consensus"
  10931..11389
  /note="LI repeat: matches 75. .529 of consensus"
  11390..11692
  /note="Alu repeat: matches 1. .301 of consensus"
  11708..12073
  /note="LI repeat: matches 516. .890 of consensus"
  12081..14318
  /note="LI repeat: matches 3060. .5390 of consensus"
  14201..15013
  /note="L1 repeat: matches 30. .872 of consensus"
  complement(16051..16359)
  /note="MER33 repeat: matches 323. .1 of consensus"
  17313..17398
  /note="MIR repeat: matches 125. .212 of consensus"
  complement(17490..18370)
  /note="L1 repeat: matches 897. .1 of consensus"
  complement(18221..20110)
  /note="LI repeat: matches 5390. .3513 of consensus"
  20133..20205
  /note="MIR repeat: matches 316. .388 of consensus"
  20154..20205
  /note="MIR repeat: matches 402. .453 of consensus"
  complement(20243..20391)
  /note="Alu repeat: matches 146. .1 of consensus"
```

```
repeat_region
  complement(20515..21058)
  /note="L1 repeat: matches 904. .356 of consensus"
  21058..21632
  /note="LI repeat: matches 3898. .4474 of consensus"
  24600..24811
  /note="MIR repeat: matches 18. .233 of consensus"
  24925..25073
  /note="MIR repeat: matches 380. .529 of consensus"
  complement(25572..25735)
  /note="MIR repeat: matches 221. .41 of consensus"
  27606..28048
  /note="MIR repeat: matches 1. .424 of consensus"
  29230..29542
  /note="L1 repeat: matches 757. .1079 of consensus"
  complement(30245..30465)
  /note="MIR repeat: matches 260. .35 of consensus"
  30903..31128
  /note="MIR repeat: matches 12. .254 of consensus"
  31940..32320
  /note="MIR repeat: matches 1. .426 of consensus"
  32748..33119
  /note="MER25 repeat: matches 1712. .2086 of consensus"
  33216..33503
  /note="LI repeat: matches 1695. .1988 of consensus"
  33779..34487
  /note="L1 repeat: matches 204. .917 of consensus"
  34896..34941
  /note="23 copies 2 mer tg 96% conserved"
  complement(34984..35711)
  /note="L1 repeat: matches 867. .101 of consensus"
  complement(35712..36011)
  /note="Alu repeat: matches 300. .1 of consensus"
  complement(36013..39594)
  /note="LI repeat: matches 5356. .1821 of consensus"
  complement(39599..40328)
  /note="LI repeat: matches 1294. .631 of consensus"
  40710..41173
  /note="match: GSS A0017146 clone 2302K13"
  41281..41417
  /note="L1 repeat: matches 919. .1055 of consensus"
  41427..42116
  /note="LI repeat: matches 2155. .2869 of consensus"
  42137..42434
  /note="Alu repeat: matches 1. .299 of consensus"
  42458..42891
  /note="LI repeat: matches 2906. .3339 of consensus"
  42909..43377
  /note="LI repeat: matches 3479. .3945 of consensus"
  complement(43397..43536)
  /note="L1 repeat: matches 675. .539 of consensus"
  43535..43777
  /note="L1 repeat: matches 673. .922 of consensus"
  44532..44707
  /note="MIR repeat: matches 1. .186 of consensus"
  44600..45088
  /note="MIR repeat: matches 69. .541 of consensus"
  complement(45124..45227)
  /note="MIR repeat: matches 488. .377 of consensus"
  complement(45442..45778)
  /note="match: GSS A0020067 clone 2303M7"
  complement(45485..45567)
  /note="MIR repeat: matches 184. .102 of consensus"
  46053..46102
  /note="25 copies 2 mer at 80% conserved"
  complement(46718..47172)
  /note="match: GSS A0004493 clone 2294C7; match: GSS A0006878 clone 2294C10"
  46787..46898
  /note="MIR repeat: matches 152. .262 of consensus"
  47199..47458
  /note="L1 repeat: matches 642. .904 of consensus"
  complement(48589..48663)
  /note="MIR repeat: matches 120. .46 of consensus"
```

```
repeat_region complement(49037..49199)
/note="L1M1 repeat: matches 661..495 of consensus"
repeat_region 49903..50335
/note="L1M1C repeat: matches 28..466 of consensus"
repeat_region complement(50441..50744)
/note="Alu10 repeat: matches 298..1 of consensus"
repeat_region 50803..50826
/note="12 copies 2 mer ta 96% conserved"
repeat_region 50844..51059
/note="L1M4A repeat: matches 834..1047 of consensus"
repeat_region 51090..51719
/note="L1 repeat: matches 4760..5390 of consensus"
repeat_region 51575..52464
/note="L1P1S repeat: matches 1..897 of consensus"
repeat_region 52842..53043
/note="L1P1S repeat: matches 697..899 of consensus"
repeat_region 53910..53978
/note="L1M2 repeat: matches 971..1044 of consensus"
repeat_region complement(53981..54055)
/note="MSTR repeat: matches 424..346 of consensus"
repeat_region complement(54054..54363)
/note="MSTR repeat: matches 310..1 of consensus"
repeat_region 54889..55140
/note="MIR repeat: matches 2..261 of consensus"
repeat_region complement(55862..55914)
/note="L1M1 repeat: matches 517..464 of consensus"
repeat_region 57446..57479
/note="17 copies 2 mer ca 88% conserved"
repeat_region 57483..57522
/note="20 copies 2 mer ga 90% conserved"
repeat_region 58284..58487
/note="L1 repeat: matches 1761..1981 of consensus"
repeat_region 58677..59040
/note="THEIC repeat: matches 3..371 of consensus"
misc_feature <58820..59202
/note="match: STS G13783"
```

```
Query Match 87.0%; Score 17.4; DB 9; Length 12325;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 2 TATTTCATGAACAAATGA 20
||||| ||||||| |||||||
Db 122021 TATTTCATGAACAAATGA 122039
```

```
RESULT 12
LOCUS HS37M17 101076 bp DNA PRI 08-AUG-1996
DEFINITION Human DNA sequence from PAC 37M17 chromosome X.
ACCESSION #278022
NID 91486355
VERSION 278022.1 GI:1486355
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
```

```
REFERENCE 1 (bases 1 to 101076)
AUTHORS Mistry,S.
JOURNAL Direct Submission
Submitted (07-AUG-1996) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RO, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequests@sanger.ac.uk
```

```
COMMENT IMPORTANT: This sequence is the entire insert of clone 37M17. The
true left end of clone 37M17 is at 1 in this sequence. The true
right end of clone 37M17 is at 101076.
37M17 is from the human PAC library.
```

```
FEATURES
source 1..101076
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
```

```
/map="X"
/clone="37M17"
3..281
/partial
/note="Alu repeat: matches 289..1 of consensus"
2344..2515
/note="MER20 element fragment"
9407..9448
/note="21 copies of 2 mer 83 % conserved"
11371..11410
/note="10 copies of 4 mer 85 % conserved"
11998..12189
/note="MIR element fragment"
12476..12775
/partial
/note="Alu repeat: matches 308..1 of consensus"
12957..13031
/note="L1 element fragment"
14836..14893
/note="MERSA element fragment"
15443..15796
/note="2 copies of 177 mer 98 % conserved"
15860..16159
/partial
/note="Alu repeat: matches 308..1 of consensus"
16169..16267
/note="MIR element fragment"
17729..18030
/partial
/note="Alu repeat: matches 308..1 of consensus"
18690..19019
/note="MSTR element fragment"
19564..19919
/note="MERIB element fragment"
19845..19919
/note="MERIA element fragment"
20108..20175
/note="17 copies of 4 mer 91 % conserved"
20205..20485
/partial
/note="Alu repeat: matches 297..1 of consensus"
21383..21449
/note="L1 element fragment"
21494..21696
/note="L1 element fragment"
21751..22083
/note="L1 element fragment"
22141..22282
/partial
/note="Alu repeat: matches 308..165 of consensus"
22283..22356
/note="L1 element fragment"
22356..22496
/partial
/note="Alu repeat: matches 162..1 of consensus"
22515..22889
/note="L1 element fragment"
22892..23000
/note="MSTR element fragment"
22907..23005
/note="MSTR element fragment"
22918..23211
/note="THEIB element fragment"
23079..23258
/note="MSTR element fragment"
23384..23602
/note="L1 element fragment"
23772..23957
/note="L1 element fragment"
24136..24249
/note="L1 element fragment"
24312..25217
/note="L1 element fragment"
```

```

repeat_region 26752..27026
/note="MER42A element fragment"
repeat_region 26776..26875
/note="L1 element fragment"
repeat_region 27831..28139
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 29106..29376
/partial
/note="Alu repeat: matches 287..1 of consensus"
repeat_region 29400..29687
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 29927..30033
/note="MLT1A element fragment"
repeat_region 29943..30005
/note="MSTC element fragment"
repeat_region 30431..30722
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 32718..32789
/note="MSTRA element fragment"
repeat_region 32733..32774
/note="TRHB1B element fragment"
repeat_region 32846..32896
/note="TRHB1B element fragment"
repeat_region 32923..33068
/note="MSTRA element fragment"
repeat_region 32961..33043
/note="TRHB1B element fragment"
repeat_region 33360..41487
/note="L1 element fragment"
repeat_region 41609..41727
/note="MLT1D element fragment"
repeat_region 42134..42220
/note="MLT1B element fragment"
repeat_region 42604..44783
/note="L1 element fragment"
repeat_region 44819..45081
/partial
/note="Alu repeat: matches 21..308 of consensus"
repeat_region 45595..45894
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 46842..47139
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 48908..49199
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 50812..50883
/note="MLT1B element fragment"
repeat_region 51397..51711
/note="MLT1B element fragment"
repeat_region 51835..52808
/note="L1 element fragment"
repeat_region 52869..52993
/partial
/note="Alu repeat: matches 8..145 of consensus"
repeat_region 52994..53107
/partial
/note="Alu repeat: matches 198..308 of consensus"
repeat_region 53138..53218
/note="L1 element fragment"
repeat_region 53226..53363
/partial
/note="Alu repeat: matches 1..144 of consensus"
repeat_region 55035..55097
/note="MLT1C element fragment"
repeat_region 55176..55292
/note="MLT1C element fragment"
repeat_region 55194..55292
/note="MLT1B element fragment"
repeat_region 56882..56944
/note="MER1B element fragment"
repeat_region 56975..57263

```

```

/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 57291..57427
/note="MER1B element fragment"
repeat_region 57354..57428
/note="MER1A element fragment"
repeat_region 57504..57783
/note="MER1B element fragment"
repeat_region 59781..60056
/partial
/note="Alu repeat: matches 306..1 of consensus"
repeat_region 60682..60973
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 63108..63158
/note="MLT2C2 element fragment"
repeat_region 63339..63377
/note="MLT2C2 element fragment"
repeat_region 64695..64823
/note="L1 element fragment"
repeat_region 65160..65612
/note="L1 element fragment"
repeat_region 65665..67096
/note="L1 element fragment"
repeat_region 67152..67235
/note="L1 element fragment"
repeat_region 67286..68295
/note="L1 element fragment"
repeat_region 68390..68571
/note="L1 element fragment"

```

```

Query Match 87.0%; Score 17.4; DB 9; Length 101076;
Best Local Similarity 94.7%; Pred. No. 2,1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TATTTCATGAAACAAATGA 20
Db 68713 TATTTCCTGAAACAAATGA 68695

RESULT 13
AF111168/
LOCUS AF111168 231464 bp DNA PRI 26-JAN-1999
DEFINITION Homo sapiens serine palmitoyl transferase, subunit II gene,
complete cds; and unknown genes.
ACCESSION AF111168
NID 94186181
VERSION AF111168.1 GI:4186181
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 231464)
AUTHORS Multimegabase Sequencing Group.
TITLE Complete sequence of the gene for serine palmitoyltransferase,
subunit II found on human chromosome 14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 231464)
AUTHORS UW Multimegabase Sequencing Group.
TITLE Direct Submision
JOURNAL Submitted (27-DEC-1998) Department of Molecular Biotechnology, Box
357730 University of Washington, Seattle, Washington 98195, USA
COMMENT On Jan 26, 1999 this sequence version replaced g1:4063505.
Sequencing methodology: high redundancy shotgun with plasmids.
Interspersed Repeats were identified with RepeatMasker (available
from http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES
source
1..231464
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q24.3-31"
/clone="BAC 50114"

```

```
repeat_region complement(250.551) /clone.lib="RBC1-11"
repeat_region /rpt_family="AluY" complement(612.683)
repeat_region /rpt_family="(TAAA)n" 722.1026
repeat_region /rpt_family="AluSg" 1027.1050
repeat_region /rpt_family="AT-rich" complement(1055.1496)
repeat_region /rpt_family="L1" complement(1680.1759)
repeat_region /rpt_family="MADE1" complement(2044.2343)
repeat_region /rpt_family="AluSg" 2345.2520
repeat_region /rpt_family="L1ME2" 2521.2819
repeat_region /rpt_family="AluY" 2820.3160
repeat_region /rpt_family="L1ME2" 3833.4128
repeat_region /rpt_family="AluSx" complement(4777.5064)
repeat_region /rpt_family="AluSx" 6726.7021
repeat_region /rpt_family="AluSx" complement(7166.7287)
repeat_region /rpt_family="FLAM_A" 7664.7966
repeat_region /rpt_family="AluSg" 7967.7994
repeat_region /rpt_family="(GAAA)n" complement(7997.8277)
repeat_region /rpt_family="MER57A" complement(8289.8569)
repeat_region /rpt_family="AluSg" complement(9087.9387)
repeat_region /rpt_family="AluSx" complement(9414.9531)
repeat_region /rpt_family="LINE2" complement(9547.9591)
repeat_region /rpt_family="(TGA)n" complement(9595.9814)
repeat_region /rpt_family="MIR" complement(10165.10287)
repeat_region /rpt_family="MIR" complement(10978.11185)
repeat_region /rpt_family="MIR" complement(11283.11475)
repeat_region /rpt_family="MIR" join(12291.12422,31490.31684,49761.49915,51955.52103,58362.58486,66381.66474,71724.71829,73351.73570,76648.76774,107289.107424,110703.110832,116467.116586)
CDS /note="intron-exon boundaries defined in relation to human cDNA sequence found in Y08686"
/codon_start=1
/product="serine palmitoyl transferase, subunit II"
/protein_id="AAD09621.1"
/db_xref="PID:94186182"
/db_xref="GI:4186182"
/translacion="MRPEGGCCCRRTVRANGCVANGEVNGYVRSSAAAAAAGG
IHHTONGGLKRPENAEETPMVAVLYGVLTFLGVLIDPFLYRTEKHHA
TEREOKDFEVLQDFENFVRLNLYMRJDNWNPICVPGARQIMEROHDYKMF
KYTGNIITKGVNMGSYNTLGFARNTGSGOEAAAYLVEYGAGVCGSTREIGNLKEE
LEELVARTIGVBAAMAYGMGFATNSMNPALVAGKCLILSEPLNHSIVLARSAGAT
IRIFKHNMSLEKLDKDAIVYGPRTRRPMKKILILVEGLYSMEGSTVRLPEVIALK
KKRAYAYLDEASHIGALGTPRGVEYFGLDPEVDVMTGTFKSGASGAYGGKK
ELIDYILRTHSHSAVAYELSPVVEOITSMKICQDGTSLGKCVQLAENRYFR
RLKEMGPIIYGNDSPPVPLMLVMPAKIGAFGRMLKRNIGVVVGVPPATPESRA
RFGLSAHTKEIILDTALKEIDVGDLQLKYSRHLVPLDRPDEITTEED"
repeat_region 12430.12532
/rpt_family="(CGGG)n"
```

```
repeat_region complement(13881.14162) /rpt_family="AluSg"
repeat_region complement(14185.14487) /rpt_family="AluY"
repeat_region complement(16941.16962) /rpt_family="AT-rich"
repeat_region complement(17032.17326) /rpt_family="AluSx"
repeat_region complement(17395.17556) /rpt_family="AluY"
repeat_region complement(17557.17858) /rpt_family="AluY"
repeat_region complement(17863.17915) /rpt_family="FLAM_A"
repeat_region 17925.17970
repeat_region /rpt_family="(TAAA)n" complement(17991.18106)
repeat_region /rpt_family="FLAM_C" 18100.18250
repeat_region /note="low quality data" complement(18108.18178)
repeat_region /rpt_family="FLAM_C" 18285.18415
repeat_region /rpt_family="FLAM_C" 18429.18729
repeat_region /rpt_family="AluSx" 18730.18979
repeat_region /rpt_family="L1MD2" complement(18991.19128)
repeat_region /rpt_family="AluY" complement(19130.19415)
repeat_region /rpt_family="AluSx" 19419.19551
repeat_region /rpt_family="L1MB5" 19738.19908
repeat_region /rpt_family="LINE2" complement(19909.19933)
repeat_region /rpt_family="(TAA)n" complement(19934.20096)
repeat_region /rpt_family="AluSg" 20281.20367
repeat_region /rpt_family="LINE2" complement(21424.21529)
repeat_region /rpt_family="MADE1" 21459.21529
repeat_region /rpt_family="MADE1" complement(21869.22165)
repeat_region /rpt_family="AluY" 22466.22496
repeat_region /rpt_family="(CA)n" complement(22524.22819)
repeat_region /rpt_family="AluSg" 23560.23858
repeat_region /rpt_family="AluY" complement(23984.24284)
repeat_region /rpt_family="AluSg" 25864.25892
repeat_region /rpt_family="AT-rich" 25943.25996
repeat_region /rpt_family="L1MC4" 25997.26324
repeat_region /rpt_family="AluY" complement(26420.26682)
repeat_region /rpt_family="LINE2" complement(26766.27294)
repeat_region /rpt_family="HERVH" complement(27297.27727)
repeat_region /rpt_family="LTR7" complement(28379.28340)
repeat_region /rpt_family="LINE2" complement(29470.29503)
repeat_region /rpt_family="AT-rich" complement(29935.30220)
```

```

repeat_region /rpt_family="Aluub"
complement(30728..31028)
repeat_region /rpt_family="AluSp"
complement(31082..31376)
repeat_region /rpt_family="AluSx"
complement(32349..32516)
repeat_region /rpt_family="AluY"
complement(32517..32814)
repeat_region /rpt_family="AluY"
complement(33486..33788)
repeat_region /rpt_family="AluSx"
complement(34256..34387)
repeat_region /rpt_family="FLAM.C"
35473..35560
repeat_region /rpt_family="Aluub"
complement(35746..36042)
repeat_region /rpt_family="AluSx"
36043..36266
repeat_region /rpt_family="LIPAL3"
36437..36728
repeat_region /rpt_family="AluSg"
36847..37137

Query Match 87.0%; Score 17.4; DB 11; Length 231464;
Best Local Similarity 94.7%; Pred. No. 26+02; 1; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1;

OY 2 TATTCATGAACAATGA 20
DB 21858 TATTCATGAACAATGA 21840

RESULT 14
CEP56H9
LOCUS CEP56H9 28291 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid F56H9, complete sequence.
ACCESSION 274473
NID 91405511
VERSION 274473.1 GI:1405511
KEYWORDS HTG; Guanine nucleotide-binding protein.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdittia; Rhabdittidae;
Rhabdittina; Rhabdittidae; Rhabdittidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 28291)
Burton, J.
Direct Submission
Submitted (20-JUN-1996) Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwhematode.wustl.edu
2 (bases 1 to 28291)
Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Haller, L., Jier, M.,
Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N.,
Latterelle, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,
Saunders, D., Showkhen, R., Smalton, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaundin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
COMMENT
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:
http://webc.sanger.ac.uk/cgi-
bin/display?db=wormbase&class=sequence&object=F56H9
Current sequence finishing criteria for the C. elegans genome

```

```

FEATURES
source
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone F56H9.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone F56H9 is at 1 in this sequence. The true
right end of clone F56H9 is at 12141 in
sequence 274476.
The true left end of clone W05B10 is at 28193 in this sequence. The
true right end of clone ZC116 is at 18114 in this sequence. The
start of this sequence (1..101) overlaps with the end of sequence
274046.
The end of this sequence (28195..28291) overlaps with the start of
sequence 274476.
location/Qualifiers
1..28291
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="F56H9"
/complement(4691..9372)
/gene="F56H9.1"
/translation(join(4691..4961,5062..5143,5193..5622,
6847..7276,7677..7818,8319..8491,8538..8619,8670..9372))
/gene="F56H9.1"
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CA98949.1"
/db_xref="PID:e1347264"
/db_xref="PID:g387738"
/db_xref="GI:387738"
/db_xref="SPTREMBL:Q20908"
/translation(join(13420..13837)
MANNYEVATINVTYTRAVSVTCQAQAVNG"
13420..13837
/gene="F56H9.2"
/translation(join(13420..13510,13755..13837)
/gene="F56H9.2"
/note="predicted using GeneFinder; cDNA EST EMBL:701752
comes from this gene"
/codon_start=1
/protein_id="CA98950.1"
/db_xref="PID:e1347265"
/db_xref="PID:g387739"
/db_xref="GI:387739"
/db_xref="SPTREMBL:Q20909"
/translation(MLEKRVLSVFVLYLFDTVATTEHHNKKDKMDYKRRKISAVGIK
DRGRIRGQFANPS"
complement(14586..16204)
/gene="gpa-8"
complement(join(14586..14651,14715..14836,14927..15080,
15128..15257,15325..15453,15503..15675,15722..15863,
15933..15975,16023..16099,16145..16204))
/gene="gpa-8"

```


EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arjan Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

FEATURES

```
source
1. 114144
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="T8n5"
/clone="T1B8"
1. 1789
/note="overlap with BAC clone PAP9
(AAC002332:150805..152593)."
misc_feature
<1..>931
/gene="T1B8.26"
/note="F4P9.44"
join(<1..153,246..281,358..425,534..594,692..>931)
/gene="T1B8.26"
join(<1..153,246..281,358..425,534..594,692..931)
/gene="T1B8.26"
/codon_start=1
/product="Barley M10 protein isolog, 5' partial"
/protein_id="AAC69142.1"
/db_xref="PID:g2253577"
/db_xref="GI:2253577"
/translation="KIQALITSDMALEIETGRNAVIGMPVWNSDRHEMSRALVHT
HIFLIFONAFETTFEWMYERGLSCFHHHALLIIRVALGVGVLCSYTLPLA
LPTOMGSTKRSEVFDOTSKALNMKNMKKSEIFGQTOPPLPILRPTGDDISAS
PANTASVDYKESDOSRDLGSP"
1770..1849
/note="exon predicted by xgrail, quality excellent"
misc_feature
2276..2349
/note="exon predicted by xgrail, quality excellent"
2403..2508
/note="exon predicted by xgrail, quality marginal"
<2660..>4843
/gene="T1B8.1"
<2660..>4843
/gene="T1B8.1"
/note="predicted by genscan and genefinder, similar to
GP14028831gnl|pid|e248504|x98130"
2660..4843
/gene="T1B8.1"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAC69139.1"
/db_xref="PID:g1707007"
/db_xref="GI:1707007"
/translation="MHPSPOTELNPHSTLIKTLKTHSQRNLVAGRAVGOIIRG
ASTCIQHANYLVNFAKCGKLAKASIFRAITICKDYVNSNLTIGSONGGISSITV
MOLFREMRADQILPNAYTLAIFKAESSLQSSVGGQALALVKNSSRFSDIVDSLV
GMYKAGAGLEDLKVFAVPERNRYTSTWAGVYATRGVEEAIVFNLFLEKEGSS
DSDYVTAIVLSLAIIYVGLROIHCITIKNGLGFVALSNALVTMYSKCSLMEAC
KMFSSGDRNSITWSAMVTGYSGNGSELAVKLFSPFSAIGIKPSFTYVGLNCS
ICYLEGKOLHSFLIKGFERHLFAFTALVDYAKAKADARKEPDICQEDVALMT
SIISGYVQNSDEBALILIRKRTACITINDPMAVSLACSSLATLELKGCHGHTI
KRGFLGVPIGSLSTMTSKCSLDEGNLVERPTPKVDVSNNAIISGLSHNGQDEA
LELFEMLAEGMEPDVTPVNIISACSHNGFERWGFYNNMSDQIGDPKVDHACM
VLLSRAGOLKEAKEFIESANIDHGLRILLSACKNGKCELYAGAEKIMALGSR
ESSTVYVSGITAGRMVDVRYKVMKHMANGVSKENGCSWELKNGYVAVFVGGTMM
PIIEFKDLYCIVSRQMIKEGVYVLDSEFVEBECTOLSTSFIT"
repeat_region
complement(4835..4868)
/rpt_family="AT-rich"
<5251..>5466
/gene="T1B8.27"
<5251..>5466
gene
```

```

CDS
/gene="T1B8.27"
/note="predicted by genscan and genefinder"
5251..5466
/gene="T1B8.27"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAC69143.1"
/db_xref="PID:g2253578"
/db_xref="GI:2253578"
/translation="MSKSEKQELPLEFSPYTKYEDIDYKKNVAGTSGHODYPGHG
GRTDAPPSGDAAPSIDSANQKAK"
6381..6427
/note="exon predicted by xgrail, quality marginal"
complement(6459..6492)
/rpt_family="(GA)n"
7124..7205
/note="exon predicted by xgrail, quality good"
join(<7252..7434,7522..7713,7864..8069,8275..9045)
/gene="T1B8.2"
<7252..9045
/gene="T1B8.2"
join(7252..7434,7522..7713,7864..8069,8275..8836)
/gene="T1B8.2"
/codon_start=1
/product="protein phosphatase 2C isolog"
/protein_id="AAC69126.1"
/db_xref="PID:g1707015"
/db_xref="GI:1707015"
/translation="MSMDVSPLLVLEGGDFNKDNTSSATEIDTLENTLDDTROISKGP
PRLHSSATRIOLANADVQCNVYMSLIDKSEFLPYRSGSCAEOGAQKQFMEDHT
CIDDLVNLGNAIOCSSGAFYQVGGDGGDAHFVAKNILRTVESSPLVYKKA
IKSARLKADYFERADSSLDISSGTTALALFIEGRLLITANAGDCAVAVGRGRATELS
KDRKPCAERYRIEKLGVVYDGLNQLAFAIGVHMKRGKSGACPLSPPELQ
ETDISEDEPFLMGCDLMDVMSQCAVTIARKETIMIHNDERGSRCEIVREALKRNIC
DNLYVIVCFEPDPFORIEIRQSRVRSISAEGNLILKGVLDGYP"
9413..9461
/note="exon predicted by xgrail, quality good"
9981..10025
/rpt_family="AT-rich"
10172..10219
/note="exon predicted by xgrail, quality marginal"
complement(10921..10977)
/rpt_family="(GAAA)n"
11621..11649
/rpt_family="AT-rich"
complement(12323..12798,13524..>13704))
/gene="T1B8.3"
complement(<12323..>13704)
/gene="T1B8.3"
complement(join(12323..12798,13524..13704))
/gene="T1B8.3"
/codon_start=1
/product="cadmium-induced protein isolog"
/protein_id="AAC69127.1"
/db_xref="PID:g1707016"
/db_xref="GI:1707016"
/translation="MHSKRPLSPESNAGNREKEKELCCSTLSEDSVDFSELTGQ
PFISSIDQSSSLTQKESNRSRQNRVORPMGKAAELRDNRKARVWLGTFDA
EEALAYDKAIFEFGRHAKLNFPEHILRVNPTOLYPSPATSHDRILVTPSPPIIAV
DILDDYGHFQSRSSDSASANLSMNLSSSSSLNHQGLRPVLEDEGVNKNISIKRRK"
13719..13745
/rpt_family="AT-rich"
14124..14156
/rpt_family="AT-rich"
complement(14317..14429)
/rpt_family="(TA)n"
complement(15330..15332)
/rpt_family="(TAA)n"
15566..15679
/note="exon predicted by xgrail, quality excellent"
15883..15983
/note="exon predicted by xgrail, quality excellent"...
```



```
misc_feature 16467..16583
              /note="exon predicted by xgrail, quality good"
              join(<16828..17139,17373..>18041)
mRNA          /gene="T1B8.29"
              <16828..>18041
gene          /gene="T1B8.29"
              /note="predicted by genscan and genefinder"
              join(16828..17139,17373..18041)
CDS           join(16828..17139,17373..18041)

Query Match      85.0%; Score 17; DB 8; Length 114144;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATTTCATGAACAAA 17
   |||||
Db 85586 GTATTTCATGAACAAA 85602
```

Search completed: September 13, 1999, 15:55:44
Job time: 4572 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:33 ; Search time 255.05 Seconds
(Without alignments)
19.619 Million cell updates/sec

Title:	US-09-325-095-22
Perfect score:	20
Sequence:	1 GATTTCATGAACAATGA 20

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36 : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	20	100.0	3412	1	051156	Human p110 cDNA, R
2	20	100.0	3207	1	051155	p110 cDNA, Recombi.
3	20	100.0	20	1	059021	Sense PCR primer f
4	20	100.0	3498	1	057012	Ptdlns 3-kinase 11
5	17.4	87.0	10670	1	T59449	Staphylococcus car
6	16.8	84.0	1586	1	Q27987	Protease from S. A
7	16.8	84.0	1558	1	Q27988	Protease from S. A
8	16.8	84.0	1019	1	V75063	Staphylococcus aut
9	15.8	79.0	255	1	T23432	Human gene signatu
10	15.8	79.0	1796	1	T29074	Arabidopsis adenyl
11	15.8	79.0	3509	1	V26289	Recombinant botuli
12	15.8	79.0	2574	1	V26280	Recombinant botuli
13	15.8	79.0	2574	1	V26282	Recombinant botuli
14	15.8	79.0	2574	1	V26283	Recombinant botuli
15	15.8	79.0	3876	1	V30579	Clostridium botuli
16	15.8	79.0	11410	1	X13137	Enterococcus faeca
17	15.8	79.0	694	1	X13663	Enterococcus faeca
18	15.4	77.0	405	1	Q60560	Human brain expres
19	15.4	77.0	7277	1	T85178	Equine rhinovirus
20	15.4	77.0	1275	1	V04010	Rat cell growth re
21	15.4	77.0	1684	1	V52391	Streptococcus pneu
22	15.4	77.0	13856	1	V74342	Staphylococcus aut
23	15.4	77.0	6723	1	X13100	Enterococcus faeca
24	15.4	77.0	2934	1	X03030	Human IL-1ra BAC c
25	15.2	76.0	2418	1	Q27886	P.falciparum GBP13
26	15.2	76.0	349	1	Q59705	Human brain expres
27	15.2	76.0	5613	1	Q63241	Crucifer 1-aminocy
28	15.2	76.0	225	1	T23910	Human gene signatu
29	15.2	76.0	447	1	V25009	H. pylori cell env
30	15.2	76.0	444	1	V24653	H. pylori ORF 03ap
31	15.2	76.0	137507	1	V119941	KSHV long unique c
32	15.2	76.0	8148	1	V52144	Streptococcus pneu
33	15.2	76.0	1588	1	T98592	DNA encoding prob
34	15.2	76.0	32207	1	V73805	KSHV LTR DNA (nuc
35	15.2	76.0	788	1	V75087	Staphylococcus aut
36	15.2	76.0	4718	1	V74610	Staphylococcus aut
37	15.2	76.0	13715	1	V74506	Staphylococcus aut
38	15.2	76.0	110000	1	X20248_03	Continuation (4 05
39	15.2	76.0	110000	1	X20248_07	Continuation (8 05
40	15.2	76.0	9848	1	X20503	Polynucleotide seq
41	15.2	75.0	612	1	X37576	Human secreted prote
42	15	75.0	110000	1	T58840_4	Continuation (5 05
43	14.8	74.0	976	1	Q25665	Sequence of the ge

c	44	45	14.8	74.0	2617	1	Q29753	DNA sequence encod
			14.8	74.0	480	1	Q31656	Pathogenesis-relat

ALIGNMENTS

```

RESULT 1
051156
ID 051156 standard; cDNA: 3412 BP.
AC 051156;
DT 12-APR-1994 (first entry)
DE Human p110 cDNA.
KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ds.
OS Human.
FH Key
FT cds
FT cds
/ *tag= a
/ note= "PI3- kinase p110"

M09321328-A.
PD 28-OCT-1993.
PF 13-APR-1993; G00761.
PR 13-APR-1993; GB-008135.
PA (LUDW-) LUDWIG INST CANCER RES.
DR Dhand R, Fxy M, Gout I, Hiles ID, Otsu M, Panayotov G,
PI Parker PJ, Vollmar S, Waterfield MD;
DR WPI; 93-351738/44.
P-PSDB; R43332.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Claim 7; Fig 16; 146bp; English.
CC Southern blot analysis was performed using a bovine cDNA probe contg.
CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
CC from a cDNA library constructed from mRNA isolated from the human
CC cell line K562. Positive clones were sequenced to give the human
CC PI3 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with PI3 kinase activity and is useful for
CC screening for (ant)agonists of PI3 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also 051155 and 057522-3.
SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 100.0%; Score 20; DB 1; Length 3412;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTCATGAACAATGCA 20
DB 3111 GTATTTCATGAACAATGCA 3130
|||||
|||||

RESULT 2
051155
ID 051155 standard; cDNA: 3207 BP.
AC 051155;
DT 12-APR-1994 (first entry)
DE p110 cDNA.
KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ds.
OS Human.
FH Key
FT cds
FT cds
/ *tag= a
/ note= "PI3- kinase p110"

M09321328-A.
PD 28-OCT-1993.
PF 13-APR-1993; G00761.
PR 13-APR-1993; GB-008135.
PA (LUDW-) LUDWIG INST CANCER RES.
DR Dhand R, Fxy M, Gout I, Hiles ID, Otsu M, Panayotov G,
PI Parker PJ, Vollmar S, Waterfield MD;
DR WPI; 93-351738/44.
P-PSDB; R43332.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Claim 7; Fig 16; 146bp; English.
CC Southern blot analysis was performed using a bovine cDNA probe contg.
CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
CC from a cDNA library constructed from mRNA isolated from the human
CC cell line K562. Positive clones were sequenced to give the human
CC PI3 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with PI3 kinase activity and is useful for
CC screening for (ant)agonists of PI3 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also 051155 and 057522-3.
SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 100.0%; Score 20; DB 1; Length 3412;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTCATGAACAATGCA 20
DB 3111 GTATTTCATGAACAATGCA 3130
|||||
|||||

RESULT 2
051155
ID 051155 standard; cDNA: 3207 BP.
AC 051155;
DT 12-APR-1994 (first entry)
DE p110 cDNA.
KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ds.
OS Human.
FH Key
FT cds
FT cds
/ *tag= a
/ note= "PI3- kinase p110"

M09321328-A.
PD 28-OCT-1993.
PF 13-APR-1993; G00761.
PR 13-APR-1993; GB-008135.
PA (LUDW-) LUDWIG INST CANCER RES.
DR Dhand R, Fxy M, Gout I, Hiles ID, Otsu M, Panayotov G,
PI Parker PJ, Vollmar S, Waterfield MD;
DR WPI; 93-351738/44.
P-PSDB; R43332.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Claim 7; Fig 16; 146bp; English.
CC Southern blot analysis was performed using a bovine cDNA probe contg.
CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
CC from a cDNA library constructed from mRNA isolated from the human
CC cell line K562. Positive clones were sequenced to give the human
CC PI3 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with PI3 kinase activity and is useful for
CC screening for (ant)agonists of PI3 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also 051155 and 057522-3.
SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 100.0%; Score 20; DB 1; Length 3412;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTCATGAACAATGCA 20
DB 3111 GTATTTCATGAACAATGCA 3130
|||||
|||||

RESULT 2
051155
ID 051155 standard; cDNA: 3207 BP.
AC 051155;
DT 12-APR-1994 (first entry)
DE p110 cDNA.
KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ds.
OS Human.
FH Key
FT cds
FT cds
/ *tag= a
/ note= "PI3- kinase p110"

M09321328-A.
PD 28-OCT-1993.
PF 13-APR-1993; G00761.
PR 13-APR-1993; GB-008135.
PA (LUDW-) LUDWIG INST CANCER RES.
DR Dhand R, Fxy M, Gout I, Hiles ID, Otsu M, Panayotov G,
PI Parker PJ, Vollmar S, Waterfield MD;
DR WPI; 93-351738/44.
P-PSDB; R43332.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Claim 7; Fig 16; 146bp; English.
CC Southern blot analysis was performed using a bovine cDNA probe contg.
CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
CC from a cDNA library constructed from mRNA isolated from the human
CC cell line K562. Positive clones were sequenced to give the human
CC PI3 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with PI3 kinase activity and is useful for
CC screening for (ant)agonists of PI3 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also 051155 and 057522-3.
SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 100.0%; Score 20; DB 1; Length 3412;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTCATGAACAATGCA 20
DB 3111 GTATTTCATGAACAATGCA 3130
|||||
|||||

RESULT 2
051155
ID 051155 standard; cDNA: 3207 BP.
AC 051155;
DT 12-APR-1994 (first entry)
DE p110 cDNA.
KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ds.
OS Human.
FH Key
FT cds
FT cds
/ *tag= a
/ note= "PI3- kinase p110"

M09321328-A.
PD 28-OCT-1993.
PF 13-APR-1993; G00761.
PR 13-APR-1993; GB-008135.
PA (LUDW-) LUDWIG INST CANCER RES.
DR Dhand R, Fxy M, Gout I, Hiles ID, Otsu M, Panayotov G,
PI Parker PJ, Vollmar S, Waterfield MD;
DR WPI; 93-351738/44.
P-PSDB; R43332.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Claim 7; Fig 16; 146bp; English.
CC Southern blot analysis was performed using a bovine cDNA probe contg.
CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
CC from a cDNA library constructed from mRNA isolated from the human
CC cell line K562. Positive clones were sequenced to give the human
CC PI3 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with PI3 kinase activity and is useful for
CC screening for (ant)agonists of PI3 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also 051155 and 057522-3.
SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 100.0%; Score 20; DB 1; Length 3412;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTCATGAACAATGCA 20
DB 3111 GTATTTCATGAACAATGCA 3130
|||||
|||||

RESULT 2
051155
ID 051155 standard; cDNA: 3207 BP.
AC 051155;
DT 12-APR-1994 (first entry)
DE p110 cDNA.
KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ds.
OS Human.
FH Key
FT cds
FT cds
/ *tag= a
/ note= "PI3- kinase p110"

M09321328-A.
PD 28-OCT-1993.
PF 13-APR-1993; G00761.
PR 13-APR-1993; GB-008135.
PA (LUDW-) LUDWIG INST CANCER RES.
DR Dhand R, Fxy M, Gout I, Hiles ID, Otsu M, Panayotov G,
PI Parker PJ, Vollmar S, Waterfield MD;
DR WPI; 93-351738/44.
```

```

FT      /note= "PI3- kinase p110"
PN      WO9321328-A.
PD      28-OCT-1993.
PF      13-APR-1993; G00761.
PR      13-APR-1992; GB-008135.
PA      (LUDM-) LUDWIG INST CANCER RES.
PI      Dhand R, Fty MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI      ParkerPJ, Volinia S, Waterfield MD;
DR      WPI: 93-351738/44.
DR      P-PSDB: R43341.
PT      Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT      activity, useful for controlling cell proliferation
CC      Claim 7; Fig 9: 146pp; English
CC      Southern blot analysis was performed using a bovine cDNA probe contg.
CC      a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
CC      from a cDNA library constructed from mRNA isolated from the human
CC      cell line KGla. Positive clones were sequenced to give the human
CC      PI3 kinase p110 sequence shown. This sequence has 95 percent
CC      homology with the bovine sequence. The domain encoding residues 19-
CC      100 of human p110 is sufficient to encode the kinase which will
CC      associate with the p85 kinase subunit. The gene may be used to
CC      provide a protein with PI3 kinase activity, and is useful for
CC      screening for (ant)agonists of PI3 kinase activity which could be
CC      useful for stimulation or inhibition of cell proliferation and hence
CC      prophylaxis or therapy. Platelet or neutrophil activity or blood
CC      glucose levels can be controlled using the kinase.
CC      See also 051156, 059012-23 and 057522-3.
SQ      Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match      100.0%; Score 20; DB 1; Length 3207;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GATTTCATGAAACAATGA 20
DB      3111 GATTTCATGAAACAATGA 3130

RESULT 3
O59021 standard; DNA: 20 BP.
O59021:
DT      12-APR-1994 (first entry)
DE      Sense PCR primer for p110.
KW      Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW      antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW      platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW      ss; amplification; p110.
CC      Synthetic.
CC      WO9321328-A.
PD      28-OCT-1993.
PR      13-APR-1992; G00761.
PR      13-APR-1992; GB-008135.
PA      (LUDM-) LUDWIG INST CANCER RES.
PI      Dhand R, Fty MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI      ParkerPJ, Volinia S, Waterfield MD;
DR      WPI: 93-351738/44.
PT      Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT      activity, useful for controlling cell proliferation
CC      Example 1; Page 41: 146pp; English.
CC      An SGBAF-1 cell line was established by transfection of bovine
CC      adrenal cortex zona fasciculata cells with psvneo. For p110
CC      poly (A)+ RNA was reversed transcribed with rTth DNA polymerase and
CC      an antisense primer. For DNA polymerisation a sense primer was added
CC      and PCR performed. The prod. was p110.
CC      See also 051155-6, 059012-23 and 057522-3.
SQ      Sequence 20 BP; 9 A; 2 C; 3 G; 6 T;

Query Match      100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      1 GATTTCATGAAACAATGA 20
DB      1 GATTTCATGAAACAATGA 20

RESULT 4
O57012 standard; cDNA to mRNA: 3498 BP.
O57012:
ID      31-AUG-1994 (first entry)
DE      Ptdins 3-kinase 110 kD catalytic subunit cDNA.
KW      110 kD catalytic subunit; phosphatidylinositol 3-kinase;
KW      transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW      Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
OS      Bos taurus.
FH      Key      Location/Qualifiers
FH      cds      1..3207
FH      /*tag= a
FH      /*product= p110

WO9403609-A.
PD      17-FEB-1994.
PR      05-AUG-1993; G01651.
PR      05-AUG-1992; GB-016654.
PA      (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PI      Goode NT, Nurse PM, Parker PJ, Waterfield MD;
DR      WPI: 94-065697/08.
DR      P-PSDB: R46294.
PT      Eukaryotic cells transformed with mammalian phospholipid or
PT      protein kinase DNA - useful in assays for compounds involved in
PT      cell growth regulation and for treating cancers
PS      Disclosure: Fig 1; 71pp; English.
CC      This sequence encodes the 110 kD catalytic subunit of the phosphatidylinositol (Ptdins) 3-kinase. This sequence was transformed into
CC      Schizosaccharomyces pombe cells under the regulatory control of the
CC      nmt promoter in an embodiment of the invention. In the presence of
CC      thiamine the promoter is inactive and the cells carrying the Ptdins
CC      catalytic subunit plasmid grow as the parental strain. In the absence
CC      of thiamine the nmt promoter functions and the Ptdins 3-kinase
CC      catalytic subunit is induced. Ptdins activity is substantially
CC      increased under these conditions. Cells containing constructs such as
CC      this, are useful in assays for detecting compounds involved in cell
CC      growth regulation. It is also used as the basis for detecting
CC      compounds for treating cancers and the formation of blood vessel
CC      plaques.
SQ      Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match      100.0%; Score 20; DB 1; Length 3498;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GATTTCATGAAACAATGA 20
DB      3111 GATTTCATGAAACAATGA 3130

RESULT 5
T99449 standard; DNA: 10670 BP.
T99449:
ID      11-MAY-1998 (first entry)
DE      Staphylococcus carnosus nitrite reductase, nitrate reductase.
KW      Staphylococcus carnosus nitrite reductase; NitrD; nitrate reductase; NARGHJ; CysG;
KW      nitrate reduction; nitrite reduction; pollutant; biotransformation;
KW      water treatment; vegetable; ds.
OS      Staphylococcus carnosus DSM 10563.
FH      Key      Location/Qualifiers
FH      -35_signal 93..98
FH      /*tag= a
FH      /*tag= b
FH      -10_signal 116..122
FH      /*tag= b

```

FT	RBS	167. .111	/*tag= c
FT	CDS	181. .2586	/*tag= d
FT		/function= nitrb gene	
FT	CDS	2589. .2903	/*tag= e
FT		/function= nird gene	
FT	CDS	2894. .3847	/*tag= f
FT		/function= cystc gene	
FT	terminator	3853. .3878	/*tag= g
FT	-35_signal	4037. .4042	/*tag= h
FT	-10_signal	4057. .4062	/*tag= i
FT	RBS	4128. .4134	/*tag= j
FT	CDS	4140. .7814	/*tag= k
FT		/function= narg gene	
FT	CDS	7804. .9381	/*tag= l
FT		/function= harh gene	
FT	CDS	9374. .9949	/*tag= m
FT		/function= nary gene	
FT	CDS	9942. .10625	/*tag= n
FT		/function= narl gene	
FT	EP-805205-A1.		
PD	05-NOV-1997.		
PF	02-MAY-1996; 201224.		
PR	02-MAY-1996; EP-201224.		
PA	(NEST) SOC PROD NESTLE SA.		
PI	Fast B, Gaier W, Goetz F, Lindgren P, Neubauer H,		
PI	Pantel I.		
DR	WPI; 97-529059/49.		
DR	P-PSDS; W26733-39.		
PT	Recombinant S. carnosus nitrate and nitrite reductase proteins -		
PT	useful for reducing nitrate and nitrite in nitrate-polluted materials		
US	Claim 3: Page 15-25; 66pp; English.		
CC	This DNA sequence from <i>Staphylococcus carnosus</i> includes genes that		
CC	code for the 2 subunits (see W26733 and W26734) of nitrite reductase		
CC	NitRD, the 3 subunits (see W26735, W26737 and W26738) of nitrate		
CC	reductase NARGH, the nitrate reductase biogenesis protein NarI		
CC	(see W26739) and the nitrite reductase <i>sirhom</i> cofactor biosynthesis		
CC	protein Cysg (see W26736). This gene region was identified		
CC	following Tn917 insertion mutagenesis studies. Claimed DNA		
CC	molecules (see T99449-51) can be used in the recombinant production		
CC	of claimed proteins (see W26733-49) involved in nitrate and nitrite		
CC	reduction in <i>S. carnosus</i> . These recombinant proteins, or cells		
CC	expressing them, in free or immobilised form, can be used in a		
CC	claimed method for reducing nitrate and nitrite in nitrate-polluted		
CC	water, vegetables and other polluted materials		
SD	Sequence 10670 BP; 3550 A; 1771 C; 2304 G; 3045 T;		

QW	Protease: PCR: Strept. ATCC12600.
KW	Protease: PCR: amplify: Staphylococcus aureus.
OS	Staphylococcus aureus.
FT	Key
FT	Location/Qualifiers
FT	352..1425
FT	/*tag- a
FT	352..555
FT	/*tag- b
FT	556..1425
FT	mat_peptide
FT	/*tag- c
PN	J04211370-A.
PD	03-AUG-1992.
PF	19-FEB-1991: 024633.
PR	20-FEB-1990: JP-040398.
PA	(SHIO.) SHIONOGI & CO LTD.
DR	WPI: 92-304938/37.
DR	P-ESD: R66842.
PT	Novel protease prep. using Bacillus or Saccharomyces host -
PT	capable of cleaving peptide bond at carboxyl terminus of glutamic
PT	acid residues in polypeptide(s)
PS	Disclosure: Page 13-15:25pp; Japanese.
CC	The sequences given in Q27987-88 encode proteases which were isolated
CC	from Staphylococcus aureus strains. The DNA sequences were isolated
CC	by PCR using the primer sequences given in Q27960-86. The protease
CC	specifically cleaves the peptide bond at the C-terminus of the
CC	glutamic acid residue in polypeptide.
SC	Sequence 1586 BP; 590 A; 302 C; 219 G; 475 T;

Query Match	84.0%	Score 16.8	DB 1	Length 1586
Best Local Similarity	90.0%	Pred. No. 40		
Matches 18	Conservative 0	Mismatches 2	Incls 0	Gaps 0
QY	1	GTATTTCATGGAACAAATGA	20	
DB	1072	GTATTTAAGAAAAAAATGA	1091	-

RESULT 7
027988
ID 027988 standard; DNA; 1558 BP.
AC 027988:
DI 11-FEB-1993 (first entry)
DE Protease from *S. aureus*.
KM Protease; PCR; amplify; *Staphylococcus*; ss.
OS *Staphylococcus aureus*.
FH Key Location/Qualifiers
FI cds 354..1364
FT /*tag- a
FT signal_peptide 354..557
FT /*tag- b
FT mat_peptide 558..1361
FT /*tag- c
FN J04213370-A.
PN 03-AUG-1992.
PD 19-FEB-1991; 024633.
PE 20-FEB-1990; JP-040398.
PR (SHTO) SHIONOGI & CO LTD.
PA WPI: 92-304938/37.
DR P-PSDB; R29644.
FT Novel protease prep. using *Bacillus* or *Saccharomyces* host -
PT capable of cleaving peptide bond at carboxyl terminus of glutamic
PI acid residues in (peptide(s))
PS Disclosure: Page 15-16; 25pp; Japanese.
CC The sequences given in J02987-88 encode proteases which were isolated
CC from *Staphylococcus aureus* strains. The DNA sequences were isolated
CC by PCR using the primer sequences given in J029960-86. The protease
CC specifically cleaves the peptide bond at the C-terminus of the
CC glutamic acid residue in polypeptide.
SQ Sequence 1558 BP; 579 A; 282 C; 221 G; 476 T;

Query Match	87.0%	Score 17.4	DB 1	Length 10670
Best Local Similarity	94.7%	Pred. No. 23		
Matches 18	Conservative	0	Mismatches 1	Indels 0
Gaps				0
OY	1	GTATTCATGAACAATG	19	
Db	6284	GTATTCATGAACAATG	6302	
RESULT	6			
Q27987				
ID	Q27987	standard	DNA	1586 BP.
AC	Q27987			
DT	-11-FEB-1993	(first entry)		

Query Match 84.0%; Score 16.8; DB 1; Length 1558;

Best Local Similarity 90.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GTATTCATGAACAATGA 20
||||| ||||| |||||
Db 1074 GTATTCATGAACAATGA 1093

RESULT 8

V75063
ID V75063 standard; DNA; 1019 BP.

AC V75063;
DT 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #752.

KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.

OS Staphylococcus aureus.

PI Key Location/Qualifiers

FT misc-feature 901..960

FT /*tag- a

FT /*note- "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence

EP-766519-A2.

PN 30-JUL-1997.

PD 07-JAN-1997; 100117.

PR 05-JAN-1996; US-009861.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,

PI Rosen CA; 97-374922/35.

DR Polynucleotide(s) and proteins derived from Staphylococcus aureus

PT stored on computer readable medium and used in the production of

PT anti-S.aureus vaccines

PS Claim 1: Page 1646-1647; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences

CC of the invention. The DNA sequences are recorded on a computer readable

CC medium, preferably selected from a floppy or hard disk, random access

CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using

CC the S.aureus DNA sequences allows putative functions to be assigned so

CC that protein-encoding or regulatory regions of commercial, therapeutic or

CC industrial importance can be obtained. Specifically, sequences which are

CC likely to encode antigens have been identified and these polypeptides can

CC be used in a vaccine composition against S.aureus infection. The

CC polypeptides can also be used in a kit for the immunodetection of

CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,

CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,

CC skin and surgical wound infections, scalded skin syndrome, toxic shock

CC syndrome, etc. Organisms transformed with the DNA sequences can be used

CC for recombinant production of the polypeptides. The new DNA sequences

CC (and their fragments) are useful as primers or probes for isolating

CC homologues of any of the S.aureus DNA sequences contained on the

CC computer readable medium.

CC Sequence 1019 BP; 364 A; 194 C; 159 G; 240 T;

Query Match 84.0%; Score 16.8; DB 1; Length 1019;
Best Local Similarity 90.0%; Pred. No. 39;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTATTCATGAACAATGA 20
||||| ||||| |||||
Db 666 GTATTCATGAACAATGA 685

RESULT 9

T23422/C
ID T23422 standard; cDNA to mRNA; 255 BP.

AC T23422;

DT 16-AUG-1996 (first entry)

DE Human gene signature HUMGS05253.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN W09514772-A1.

PD 01-JUN-1995.

PE 11-NOV-1994; J01916.

PF 12-NOV-1993; JP-355504.

PA (MATS/) MATSUBARA K.

PI (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

DR WPI, 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS Claim 1: Page 1365; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

CC Sequence 255 BP; 84 A; 39 C; 31 G; 95 T;

Query Match 79.0%; Score 15.8; DB 1; Length 255;

Best Local Similarity 89.5%; Pred. No. 1e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TATTCATGAACAATGA 20
||||| ||||| |||||
Db 78 TATTCATGAACAATGA 60

RESULT 10

T29074
ID T29074 standard; cDNA; 1796 BP.

AC T29074;

DT 13-SEP-1996 (first entry)

DE Arabidopsis adenylosuccinate lyase cDNA ADSL-1.

KW Adenylosuccinate lyase; ADSL; inhibitor; herbicide; ss.

OS Arabidopsis thaliana.

FS Key Location/Qualifiers

FT cds 14..1603

FT /*tag- a

FT misc-difference 1652

FT /*tag- b

FT /*note- "base 1652 is given as 'y' in the

FT specification"

PN W09618739-A1.

PD 20-JUN-1996.

PE 06-DEC-1995; E04796.

PF 14-DEC-1994; US-355770.

PR 26-OCT-1995; US-548509.

PA (CIBA) CIBA GEIGY AG.

PI Guyer CD, Ward ER;

DR WPI, 96-300656/30.

P-PDB; R97547.

PT DNA mol. encoding plant adenylosuccinate lyase (ADSL) - useful to

PT screen for inhibitors of ADSL, which may be used as herbicides

PS Claim 4: Page 13-17; 37pp; English.

CC A full-length cDNA (T29074), designated ADSL-1, codes for

CC adenylsuccinate lyase (ADSL) (R97547), an enzyme useful for
CC screening ADSL-inhibitors or potential use as herbicides. It was
CC isolated from an Arabidopsis thaliana lambda ZAP library using a cDNA
CC clone isolated by functional complementation of the E. coli purB
CC mutant JX268. A second cDNA fragment (T31801), ADSL-2, was also
CC isolated and represents a second distinct gene (72% sequence homology
CC to ADSL-1). ADSL-1 cDNA can be incorporated into a vector and used
CC for predn. of recombinant ADSL, or used as a probe to isolate
CC homologous genes, for gene mapping or for diagnosis of plant diseases
CC associated with ADSL.
SO Sequence 1796 BP; 531 A; 336 C; 402 G; 526 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1796;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTATTCATGAACAATG 19
DB 502 GTATTCATGAACAATG 520

RESULT 11

V26289
ID V26289 standard; DNA: 3509 BP.
AC V26289; (first entry)
DE Recombinant botulinum neurotoxin type B LH728/B encoding DNA.
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KM immunogen; detection; tetanus; non-toxic; toxin; ds.
OS Synthetic.
FH Clostridium botulinum.
FT Key Location/Qualifiers
FT CDS 1..3509
FT /*tag= a
FT /product= "LH728/B"
FT /note= "no stop codon given"

PN MO9807864-A1.
PD 26-FEB-1998.
PF 22-AUG-1997; G02273.
PR 13-DEC-1996; GB-025996.
PR 23-AUG-1996; GB-017671.
PA (MTCR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SEPEX-) SPEYWOOD LAB LTD.
PI Foster KA, Quinn CP, Shone CC;
PI WPI: 98-169168/15.
DR P-PSDB: W56017.
DR

PT Recombinant neurotoxin polypeptides - used to develop therapeutic
PT agents, immunogens or as non-toxic standards for the detection of
PT neurotoxins
PS Disclosure: Page 87-91; 137pp; English.
CC The present sequence encodes a recombinant neurotoxin protein from
CC +he present invention. The present invention describes recombinant
CC neurotoxin proteins which comprise a first and second domain, where
CC the first domain is adapted to cleave one or more vesicle or
CC plasma-membrane associated proteins essential to exocytosis, and where
CC the second domain is adapted: (a) to translocate the protein into a
CC cell; (b) to increase the solubility of the protein compared to the
CC solubility of the first domain on its own, or (c) both to translocate
CC the protein into a cell and to increase the solubility of the protein
CC compared to the solubility of the first domain on its own, the protein
CC being free of clostridial neurotoxin (CN) and free of CN precursor that
CC can be converted into toxin by proteolytic action. The recombinant
CC proteins can be used as therapeutic agents for targeting cells
CC expressing a relevant substrate. The products can also be used as
CC immunogens and as non-toxic standards for the assessment and development
CC of in vitro assays for the detection of functional botulinum or tetanus
CC neurotoxins either in foodstuffs or in environmental samples.
SO Sequence 3509 BP; 1468 A; 340 C; 548 G; 1153 T;

Query Match 79.0%; Score 15.8; DB 1; Length 3509;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TATTCATGAACAATGA 20
DB 2344 TATTCATGAACAATGA 2362

RESULT 12

V26290
ID V26290 standard; DNA: 2574 BP.
AC V26290; (first entry)
DE Recombinant botulinum neurotoxin type B LH417/B encoding DNA.
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KM immunogen; detection; tetanus; non-toxic; toxin; ds.
OS Synthetic.
FH Clostridium botulinum.
FT Key Location/Qualifiers
FT CDS 1..2574
FT /*tag= a
FT /product= "LH417/B"
FT /note= "no stop codon given"

PN MO9807864-A1.
PD 26-FEB-1998.
PF 22-AUG-1997; G02273.
PR 13-DEC-1996; GB-025996.
PR 23-AUG-1996; GB-017671.
PA (MTCR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SEPEX-) SPEYWOOD LAB LTD.
PI Foster KA, Quinn CP, Shone CC;
PI WPI: 98-169168/15.
DR P-PSDB: W56018.
DR

PT Recombinant neurotoxin polypeptides - used to develop therapeutic
PT agents, immunogens or as non-toxic standards for the detection of
PT neurotoxins
PS Disclosure: Page 95-98; 137pp; English.
CC The present sequence encodes a recombinant neurotoxin protein from
CC +he present invention. The present invention describes recombinant
CC neurotoxin proteins which comprise a first and second domain, where
CC the first domain is adapted to cleave one or more vesicle or
CC plasma-membrane associated proteins essential to exocytosis, and where
CC the second domain is adapted: (a) to translocate the protein into a
CC cell; (b) to increase the solubility of the protein compared to the
CC solubility of the first domain on its own, or (c) both to translocate
CC the protein into a cell and to increase the solubility of the protein
CC compared to the solubility of the first domain on its own, the protein
CC being free of clostridial neurotoxin (CN) and free of CN precursor that
CC can be converted into toxin by proteolytic action. The recombinant
CC proteins can be used as therapeutic agents for targeting cells
CC expressing a relevant substrate. The products can also be used as
CC immunogens and as non-toxic standards for the assessment and development
CC of in vitro assays for the detection of functional botulinum or tetanus
CC neurotoxins either in foodstuffs or in environmental samples.
SO Sequence 2574 BP; 1060 A; 273 C; 403 G; 838 T;

Query Match 79.0%; Score 15.8; DB 1; Length 2574;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TATTCATGAACAATGA 20
DB 2344 TATTCATGAACAATGA 2362

RESULT 13

V26292
ID V26292 standard; DNA: 2574 BP.
AC V26292; (first entry)
DE Recombinant botulinum neurotoxin type B LH417/B E.coli codon bias DNA.
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KM immunogen; detection; tetanus; non-toxic; toxin; ds.

OS Synthetic.
 OS Clostridium botulinum.
 PN W09807864-A1.
 PD 26-FEB-1998.
 PF 22-AUG-1997: G02273.
 PR 13-DEC-1996: GB-025996.
 PR 23-AUG-1996: GB-017671.
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
 PI (SPEY-) SPEYWOOD LAB LTD.
 PI Foster KA, Quinn CP, Shone CC;
 DR WPI: 98-169168/15.
 PT Recombinant neurotoxin polypeptides - used to develop therapeutic
 CC agents, immunogens or as non-toxic standards for the detection of
 CC neurotoxins.
 CC Example 2: Page 111-112: 137pp; English.
 CC The present sequence encodes LH417/B, a recombinant neurotoxin
 CC protein, with an Escherichia coli codon bias, from the present
 CC invention. The present invention describes recombinant neurotoxin
 CC proteins which comprise a first and second domain, where the first
 CC domain is adapted to cleave one or more vesicle or plasma-membrane
 CC associated proteins essential to exocytosis, and where the second
 CC domain is adapted: (a) to translocate the protein into a cell; (b)
 CC to increase the solubility of the protein compared to the solubility
 CC of the first domain on its own, or (c) both to translocate the
 CC protein into a cell and to increase the solubility of the protein
 CC compared to the solubility of the first domain on its own, the protein
 CC being free of clostridial neurotoxin (CN) and free of CN precursor that
 CC can be converted into toxin by proteolytic action. The recombinant
 CC proteins can be used as therapeutic agents for targeting cells
 CC expressing a relevant substrate. The products can also be used as
 CC immunogens and as non-toxic standards for the assessment and development
 CC of in vitro assays for the detection of functional botulinum or tetanus
 CC neurotoxins either in foodstuffs or in environmental samples.
 CC Sequence 2574 BP; 1038 A; 319 C; 412 G; 805 T;

Query Match 79.0%; Score 15.8; DB 1; Length 2574;
 Best Local Similarity 89.5%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0Y 2 TATTCATGAAACAATGA 20
 DB 2344 TATTTAATGAAAAAATGA 2362

RESULT 14
 V26293
 ID V26293 standard; DNA: 2574 BP.
 AC V26293:
 DT 27-JUL-1998 (first entry)
 DE Recombinant botulinum neurotoxin type B LH417/B encoding DNA.
 KM Botulinum; recombinant; Clostridium botulinum; neurotoxin;
 KW Immunogen; detection; tetanus; non-toxic; toxin; ds.
 OS Synthetic.
 OS Clostridium botulinum.
 PN W09807864-A1.
 PD 26-FEB-1998.
 PF 22-AUG-1997: G02273.
 PR 13-DEC-1996: GB-025996.
 PR 23-AUG-1996: GB-017671.
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
 PI (SPEY-) SPEYWOOD LAB LTD.
 PI Foster KA, Quinn CP, Shone CC;
 DR WPI: 98-169168/15.
 PT Recombinant neurotoxin polypeptides - used to develop therapeutic
 CC agents, immunogens or as non-toxic standards for the detection of
 CC neurotoxins.
 CC Example 2: Page 112-114: 137pp; English.
 CC The present sequence encodes LH417/B, a recombinant neurotoxin
 CC protein, where bases 691-1641 of the native BONT/B sequence have
 CC been replaced by a degenerate DNA coding for amino acid residues
 CC 231-547 of the native BONT/B protein, from the present
 CC invention. The present invention describes recombinant neurotoxin

CC proteins which comprise a first and second domain, where the first
 CC domain is adapted to cleave one or more vesicle or plasma-membrane
 CC associated proteins essential to exocytosis, and where the second
 CC domain is adapted: (a) to translocate the protein into a cell; (b)
 CC to increase the solubility of the protein compared to the solubility
 CC of the first domain on its own, or (c) both to translocate the
 CC protein into a cell and to increase the solubility of the protein
 CC compared to the solubility of the first domain on its own, the protein
 CC being free of clostridial neurotoxin (CN) and free of CN precursor that
 CC can be converted into toxin by proteolytic action. The recombinant
 CC proteins can be used as therapeutic agents for targeting cells
 CC expressing a relevant substrate. The products can also be used as
 CC immunogens and as non-toxic standards for the assessment and development
 CC of in vitro assays for the detection of functional botulinum or tetanus
 CC neurotoxins either in foodstuffs or in environmental samples.
 CC Sequence 2574 BP; 1001 A; 466 C; 422 G; 685 T;

Query Match 79.0%; Score 15.8; DB 1; Length 2574;
 Best Local Similarity 89.5%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0Y 2 TATTCATGAAACAATGA 20
 DB 2344 TATTTAATGAAAAAATGA 2362

RESULT 15
 V30579
 ID V30579 standard; DNA: 3876 BP.
 AC V30579:
 DT 07-DEC-1998 (first entry)
 DE Clostridium botulinum type B toxin gene from Danish strain.
 KM Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
 KW botulism; BotB; ds.
 OS Clostridium botulinum serotype B Danish strain.
 PN W09808640-A1.
 PD 05-MAR-1998.
 PF 28-AUG-1997: U15394.
 PR 28-AUG-1996: US-704159.
 PA (OPHI-) OPHIDIAN PHARM INC.
 PI Thallely BS, Williams JA.
 DR WPI: 98-230234/20.
 DR P-PSDB: W68392.
 PT Host cell containing recombinant expression vector encoding
 CC Clostridium botulinum type B or E toxin - useful to treat humans
 CC and other animals at risk of intoxication with clostridial toxin
 CC Example 35: Page 291-296: 428pp; English.
 CC This is the coding region of the Clostridium botulinum serotype B
 CC (Danish strain) toxin gene that codes for a 1291-amino acid
 CC polypeptide (see W68392). The C fragment (see W68394) of the
 CC B toxin has been expressed as histidine-tagged protein in Escherichia
 CC coli host cells. The invention relates to C. botulinum recombinant
 CC toxin polypeptides. Methods are provided which allow for the
 CC isolation of soluble recombinant proteins free of significant
 CC endotoxin contamination. Preferred hosts for production of the
 CC recombinant proteins are E. coli, insect cells and yeast cells.
 CC The recombinant proteins are used as immunogens for the production
 CC of vaccines and antitoxins that are useful in the treatment of
 CC humans and animals at risk of intoxication with clostridial toxin.
 CC Sequence 3876 BP; 1612 A; 370 C; 617 G; 1277 T;

Query Match 79.0%; Score 15.8; DB 1; Length 3876;
 Best Local Similarity 89.5%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0Y 2 TATTCATGAAACAATGA 20
 DB 2344 TATTTAATGAAAAAATGA 2362

Thu Sep 16 12:23:39 1999

us-09-325-095-22.rng

Page 7

Search completed: September 13, 1999, 15:59:34
Job time: 4801 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:56 ; Search time 1694.61 Seconds

(without alignments)
23.280 Million cell updates/sec

Title: US-09-325-095-22

Perfect score: 20
Sequence: 1 GATTTCATGACAAATGA 20

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est23:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB	ID	Description
1	20	292	23	D56232	D56232 HUM420G03B
2	20	264	24	N23534	N23534 YV91G08.r1
3	20	475	25	N44819	N44819 YX39A10.r1
4	20	433	26	W76227	W76227 Z64C12.r1
5	20	343	31	AA296585	AA296585 EST114216
6	20	633	49	A1632172	A1632172 t885907.x
7	19	241	25	N44835	N44835 YX39A10.r1
8	17.4	87.0	367	AA499534	AA499534 v193h08.r
9	17.4	285	49	A1643250	A1643250 v193h08.y
10	16.8	472	30	AA242876	AA242876 z164d01.x
11	16.8	360	36	C65080	C65080 C65080.f
12	16.8	398	42	A1099642	A1099642 34368 Lam
13	16.4	264	21	F09317	F09317 HSC22H112 n
14	16.4	432	22	R44353	R44353 Y936B02.s1
15	16.4	358	23	R87639	R87639 YP89C01.s1
16	16.4	352	24	H82513	H82513 Y869B12.s1
17	16.4	321	24	H94229	H94229 YV07F01.r1
18	16.4	355	24	N21852	N21852 SMNHADA0180
19	16.4	291	27	C05875	C05875 C05875.Huma
20	16.4	730	34	C24336	C24336 C24336 Dict
21	16.4	649	34	C24406	C24406 C24406 Dict
22	16.4	343	36	AA632828	AA632828 np84d05.s
23	16.4	481	38	AA796325	AA796325 vs94e06.f
24	16.4	475	38	AA796972	AA796972 vs94e07.f
25	16.4	293	43	A1168589	A1168589 cx11d04.s
26	16.4	610	44	A0038541	A0038541 A0038541
27	16.4	420	46	A1428066	A1428066 mm38b08.x
28	16.4	1823	50	AF013092	AF013092 AF013092
29	16	489	44	A1283067	A1283067 qm62c11.x
30	16	556	51	A1727149	A1727149 BNLGH1745
31	16	650	51	A1727527	A1727527 BNLGH1832
32	16	743	51	A1729279	A1729279 BNLGH1130
33	15.8	360	20	D36276	D36276 CELK030H7F
34	15.8	294	20	T07560	T07560 EST05450.Fe
35	15.8	324	20	T16380	T16380 NIB124 Nor
36	15.8	248	20	T17087	T17087 NIB314 Norm
37	15.8	331	20	T34362	T34362 EST6619 Hu
38	15.8	459	20	T51830	T51830 YB54C02.s1
39	15.8	275	20	Z41154	Z41154 HSC22P092 n
40	15.8	79.0	448	H05472	H05472 Y181F06.s1
41	15.8	79.0	413	H13474	H13474 Y115C06.r1
42	15.8	79.0	331	R37868	R37868 YC98F08.s1
43	15.8	231	22	R42218	R42218 YF90A11.s1
44	15.8	363	22	R49580	R49580 Y668F11.s1
45	15.8	375	51	A1716960	A1716960 UT-R-Y0-a

ALIGNMENTS

RESULT 1
D56232/c D56232 292 bp mRNA 31-AUG-1995
LOCUS HUM420G03B Clontech human fetal brain polyA+ mRNA (#6535) Homo
DEFINITION sapiens cDNA clone GEN-420G03 5', mRNA sequence.
ACCESSION D56232
NID 9970756
VERSION D56232.1 GI:970756

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 292)

TITLE
JOURNAL Fujitara, T., Hirano, H., Katagiri, T., Kawai, A., Kuge, Y., Nagata, M.,
Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shimomiyu, H.,
Takahashi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y.,
Maekawa, H., Shin, S. and Nakamura, Y.
Fujitara et al. (1995)
Unpublished (1995)
On Sep 21, 1992 this sequence version replaced gi:276155.

COMMENT

FEATURES
source
Contact: Tsutomu Fujitara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-85-2888
Fax: 0886-85-2888
Insert Length: 770 Std Error: 0.00
High quality sequence stop: 331.
Location/Qualifiers
1..292
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="Clontech human fetal brain polyA+ mRNA
(#6535)"

BASE COUNT
ORIGIN 78 a 59 c 55 g 100 t

Query Match 100.0%; Score 20; DB 23; Length 292;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 GTATTTCATGAAACAATGA 20
DB 116 GTATTTCATGAAACAATGA 97
|||||

RESULT 2
LOCUS N23534 264 bp mRNA EST 28-DEC-1995
DEFINITION yv91908.r1 Soares melanocyte 2NBHM Homo sapiens cDNA clone
IMAGE:250142 5' similar to SP:P100_BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT ;, mRNA sequence.
ACCESSION N23534
NID G1137684
VERSION N23534.1 GI:1137684
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 264)
REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The Washu-Merck EST Project
Unpublished (1995)
On Apr 14, 1993 this sequence version replaced gi:692929.

TITLE
JOURNAL
COMMENT

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
High quality sequence stops: 225

Source: IMAGE Consortium, LNL.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1135 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 225.
Location/Qualifiers
1..264
/organism="Homo sapiens"
/db_xref="GDB:3867848"
/db_xref="taxon:9606"
/map="14"
/clone_id="IMAGE:250142"
/clone_id="Soares melanocyte 2NBHM"
/issue_type="melanocyte"
/sex="Male"
/lab_host="PH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [3'
TGTTACCAATCTGAGTGGAGCGCCGAGTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT
ORIGIN 94 a 53 c 52 g 60 t

Query Match 100.0%; Score 20; DB 24; Length 264;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 GTATTTCATGAAACAATGA 20
DB 91 GTATTTCATGAAACAATGA 110
|||||

RESULT 3
LOCUS N44819 475 bp mRNA EST 13-FEB-1996
DEFINITION yv39a10.r1 Soares melanocyte 2NBHM Homo sapiens cDNA clone
IMAGE:273594 5' similar to SW:P114_BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;,
mRNA sequence.
ACCESSION N44819
NID G1185985
VERSION N44819.1 GI:1185985
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 475)
REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The Washu-Merck EST Project
Unpublished (1995)
On May 8, 1993 this sequence version replaced gi:800198.

TITLE
JOURNAL
COMMENT

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: T7
High quality sequence stop: 402.
Location/Qualifiers
1..475

FEATURES
Source
/organism="Homo sapiens"
/db_xref="GDB:3883236"
/db_xref="taxon:9606"
/clone="IMAGE:273594"
/clone_id="Soares melanocyte 2NDHM"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'] TGTACCAATCTGAGTGGAGCGCCGACATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 170 a 90 c 85 g 128 t 2 others
ORIGIN

Query Match 100.0%; Score 20; DB 25; Length 475;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATTTCATGAACAAATGA 20
|||||
DB 152 GTATTTCATGAACAAATGA 171

RESULT 4
LOCUS W76227 433 bp mRNA EST 16-OCT-1996
DEFINITION z664c12.1 Soares_fetal_heart_NbH19W Homo sapiens cDNA clone IMAGE:345430 5' similar to SW:P11A_BOVIN P32871 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ; mRNA sequence.
W76227
ACCESSION W76227.1 GI:1386490
VERSION g1386490
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 433)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maira,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisaki,E., Waterston,R., Williamson,A., Womlmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
On Apr 14, 1993 this sequence version replaced g1:693546.

TITLE JOURNAL
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1049 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 391.
Location/Qualifiers
1..433

/organism="Homo sapiens"
/db_xref="GDB:1270805"
/db_xref="taxon:9606"
/map="20"
/clone="IMAGE:345430"
/clone_id="Soares_fetal_heart_NbH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'] TGTACCAATCTGAGTGGAGCGCCGACATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbH19W."

BASE COUNT 148 a 78 c 90 g 116 t 1 others
ORIGIN

Query Match 100.0%; Score 20; DB 26; Length 433;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATTTCATGAACAAATGA 20
|||||
DB 276 GTATTTCATGAACAAATGA 295

RESULT 5
LOCUS AA298585 343 bp mRNA EST 18-APR-1997
DEFINITION EST114216 HSC172 cells II Homo sapiens cDNA 5' end similar to phosphatidylinositol 3-kinase, alpha catalytic subunit, mRNA sequence.
AA298585
ACCESSION AA298585.1 GI:1950938
VERSION g1950938
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 343)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Mai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,R.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Lin,L.-I., Marmaros,D.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y., Bedarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dlake,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,M.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,D., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wel,Y.F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Pannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
On Sep 13, 1996 this sequence version replaced g1:1392803.
Other ESTs: THC168479
Contact: Kerlavage, AR

TITLE JOURNAL
MEDLINE
COMMENT
On Sep 13, 1996 this sequence version replaced g1:1392803.
Other ESTs: THC168479
Contact: Kerlavage, AR

Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavet@igrr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/cdb/hg1/hg1.html>)
Seq primer: M13 Reverse.

FEATURES

source

1. 343
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="taxon:9606"
/clone_lib="HSC172 cells T1"
/cell_type="fibroblast"
/cell_line="HSC172 (60PDL)"
/dev_stage="fetal"
/note="Organ: lung; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 132 a 61 c 64 g 84 t 2 others
ORIGIN

Query Match 100.0%; Score 20; DB 31; Length 343;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTATTTCATGAACAATGA 20
|||||
Db 67 GTATTTCATGAACAATGA 86

RESULT 6
A1632172 633 bp mRNA EST 26-APR-1999
LOCUS ts85g07.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:228108 3'
DEFINITION similar to SW:P1LA_BOVIN P32871 PHOSPHATIDYLINOSITOL 3-KINASE
CATALYTIC SUBUNIT, ALPHA ISOFORM ;, mRNA sequence.
ACCESSION A1632172
NID 94683502
VERSION A1632172.1 GI:4663502
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 633)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL On May 7, 1998 this sequence version replaced gi:3121411.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

FEATURES

source

Seq primer: -40UP from Gibco
High quality sequence stop: 446.
Location/Qualifiers
1. 633
/organism="Homo sapiens"

/db_xref="taxon:9606"
/map="15q26.1"
/clone_image="228108"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker; Plasmid DNA from the normalized library
NCI_CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (clones 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldi."
BASE COUNT 169 a 109 c 108 g 246 t
ORIGIN

Query Match 100.0%; Score 20; DB 49; Length 633;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTATTTCATGAACAATGA 20
|||||
Db 440 GTATTTCATGAACAATGA 421

RESULT 7
M44835 241 bp mRNA EST 13-FEB-1996
LOCUS y399e10.r1 Soares melanocyte 2NDHM Homo sapiens cDNA clone
IMAGE:273642 5', similar to SW:P1LA_BOVIN P32871
DEFINITION PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;,
mRNA sequence.
ACCESSION M44835
NID 91186001
VERSION M44835.1 GI:1186001
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 241)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Roffling, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
JOURNAL On May 8, 1995 this sequence version replaced gi:800214.
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: 77
High quality sequence stop: 1.
Location/Qualifiers
1. 241
/organism="Homo sapiens"
/db_xref="GDB:3883284"
/db_xref="taxon:9606"
/map="3 p24.2-p21.1"
/clone_image="IMAGE:273642"
/clone_lib="Soares melanocyte 2NDHM"

FEATURES

source

/sex="Male"
/tissue-type="melanocyte"
/lab-host="DH108 (ampicillin resistant)"
/note="vector: p773D (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer (5'
TGTTCACATCGATGAGTGGAGCGCCGACAGTTTTTTTTTTTTTTT 3')
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 69 a 48 c 44 g 56 t 24 others

ORIGIN

Query Match 95.0%; Score 19; DB 25; Length 241;
Best Local Similarity 95.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTATTTCATGAACAATGA 20
|||||
Db 148 GTATTCATGAACAATGA 167

RESULT 8
AA499534 367 bp mRNA EST 01-JUL-1997
LOCUS *v193h08.y1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
DEFINITION IMAGE:919839 5', mRNA sequence.
ACCESSION AA499534
NID 92234501
VERSION AA499534.1 GI:2234501
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 367)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lemon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMT Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1392797.

Contact: Marra M/Mouse EST Project
WashU-HMT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:532055
Seq primer: -28m3 rev1 ET from Amerham
High quality sequence stop: 345.

FEATURES
source
Location/Qualifiers
1..367
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone IMAGE:919839"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue-type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site_1:
/note="Organ: heart; Vector: pBluescript SK-; Site_1:

EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTGTGTGTGTGTGT 3'."

BASE COUNT 100 a 86 c 98 g 83 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 34; Length 367;
Best Local Similarity 94.7%; Pred. No. 5.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TATTTCATGAACAATGA 20
|||||
Db 154 TATTCATGAACAATGA 172

RESULT 9
A1643250 285 bp mRNA EST 29-APR-1999
LOCUS *v193h08.y1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
DEFINITION IMAGE:919839 5', mRNA sequence.
ACCESSION A1643250
NID 94721725
VERSION A1643250.1 GI:4721725
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 285)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,T.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schuck,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,T., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980088.

Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:532055
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40RP from G1bco
High quality sequence stop: 272.

FEATURES
source
Location/Qualifiers
1..285
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/map="7"
/clone IMAGE:919839"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue-type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTGTGTGTGTGT 3'."

BASE COUNT 84 a 59 c 76 g 66 t
 CACCTG

Query Match 87.0%; Score 17.4; DB 49; Length 285;
 Best Local Similarity 94.7%; Pred. No. 6.1e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TATTCATGAAACAATGA 20
 ||| ||||| ||||| |||||
 Db 154 TATTCATGAAACAATGA 172

RESULT 10
 AA242876 472 bp mRNA EST 06-AUG-1997
 LOCUS zrf6401.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:668161
 DEFINITION 5', mRNA sequence.
 ACCESSION AA242876
 NID 91873669
 VERSION AA242876.1 GI:1873669
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Creniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 472)
 Hillier, L., Allen, M., Bowles, L., Dubugue, T., Gelsel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
 Moore, B., Schellenberg, K., Stepec, M., Tan, F., Theising, B.,
 White, Y., Wylie, T., Waterston, R. and Wilson, R.
 WashU-Merck EST Project 1997
 *Unpublished (1997)
 JOURNAL On May 5, 1995 this sequence version replaced gi:798270.
 COMMENT

TITLE
 JOURNAL
 COMMENT
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 761 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 460.
 location/Qualifiers

FEATURES
 source
 1..472

/organism="Homo sapiens"
 /db_xref="GDB:5562128"
 /db_xref="taxon:9606"
 /map="11"
 /clone="IMAGE:668161"
 /clone_lib="Soares_NhHMPu_S1"
 /tissue_type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10b"
 /note="Organ: mixed (see below); Vector: pT73D-Pac
 (Pharmacia) with a modified polylinker; Site 1: Not I;
 Site 2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NHM, pregnant uterus
 NHMPu, and fetal heart NBH11W) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of I.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."

BASE COUNT 137 a 91 c 87 g 157 t
 ORIGIN

Query Match 84.0%; Score 16.8; DB 30; Length 472;

Best Local Similarity 90.0%; Pred. No. 8.4e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTATTCATGAAACAATGA 20
 ||||| ||||| ||||| |||||
 Db 127 GTATTCATGAAACAATGA 146

RESULT 11
 C65080 360 bp mRNA EST 22-SEP-1997
 LOCUS C65080/c
 DEFINITION C65080 Yuii Kohara unpublished cDNA Caenorhabditis elegans cDNA
 clone yk39898 5', mRNA sequence.
 ACCESSION C65080
 NID 92423785
 VERSION C65080.1 GI:2423785
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabdilitia;
 Rhabdilitia; Rhabdilitidae; Rhabdilitidae; Pelodierinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 360)
 Kohara, Y., Mutohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
 Sano, M., Miyata, A. and Nishigaki, A.
 Expression map of the C.elegans genome
 Unpublished (1996)
 JOURNAL On Sep 12, 1996 this sequence version replaced gi:1402204.
 COMMENT

TITLE
 JOURNAL
 COMMENT
 Contact: Yuii Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 location/Qualifiers

FEATURES
 source
 1..360

/organism="Caenorhabditis elegans"
 /strain="CB1489 hlm-8(e1489)"
 /db_xref="taxon:6239"
 /clone="yk39898"
 /clone_lib="Yuii Kohara unpublished cDNA"
 /note="dev stage=varied, sex=Hermaphrodite male,
 tissue_type=whole animal"

BASE COUNT 100 a 69 c 62 g 125 t 4 others
 ORIGIN

Query Match 84.0%; Score 16.8; DB 36; Length 360;
 Best Local Similarity 90.0%; Pred. No. 9.5e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTATTCATGAAACAATGA 20
 ||||| ||||| ||||| |||||
 Db 183 GTCTTCTTCGAAACAATGA 164

RESULT 12
 A1099642 398 bp mRNA EST 21-AUG-1998
 LOCUS A1099642/c
 DEFINITION A1099642 Lambda-PRL2 Arabidopsis thaliana cDNA clone 1607XP 3', mRNA
 sequence.
 ACCESSION A1099642
 NID 93449954
 VERSION A1099642.1 GI:3449954
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 REFERENCE 1 (bases 1 to 398)

AUTHORS Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomasnow, M., Retzel, E., and Somerville, C.

TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones

JOURNAL MEDLINE Plant Physiol. 106, 1241-1255 (1994)

COMMENT On Jan 14, 1998 this sequence version replaced gi:1798073.

CONTACT: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 2231jcn@pml.c1.msu.edu

The sequence entry for this EST has been reverse complemented and is being submitted in the sense orientation.

Seq primer: M13-21.

FEATURES
Location/Qualifiers
1..398
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone="160F7XP"
/note="Vector: lambda Z1P-Lox; Site_1: Sal; Site_2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques. The vector is BRU's lambda Z1P-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."

BASE COUNT 100 a 102 c 70 g 123 t 3 others

ORIGIN

Query Match 84.0%; Score 16.8; DB 42; Length 398;
Best Local Similarity 90.0%; Pred. No. 9e+02; Mismatches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTATTTCATGAACAATGA 20
|||||
Db 305 GTATTTCATTAATAATGA 286

RESULT 13
F09317/c 264 bp mRNA EST 23-FEB-1995
LOCUS HSC22H12 normalized infant brain cDNA Homo sapiens cDNA clone
DEFINITION C-22h11 3', mRNA sequence.
ACCESSION F09317
NID 9678473
VERSION F09317.1 GI:678473
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 264)
Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Marigoe-Samson, R., Pietu, G., Pouliot, Y., Sebastien, K., Kachich, C., and Tessier, A.
TITLE IMAGE: molecular integration of the analysis of the human genome and its expression
JOURNAL MEDLINE C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
COMMENT 95277534

CONTACT: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr

Single read, removed at sequence 5' end
Genexpress_library_id: C; Genexpress_sequence_id: alc-22h11
Insert Length: 528 Std Error: 0.00
Seq primer: (-21)M13 universal
High quality sequence stop: 1.

FEATURES
Location/Qualifiers
1..264
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="C-22h11"
/clone.lib="normalized infant brain cDNA"
/sex="female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lambda Z1P-Lox; Site_1: HindIII; Site_2: NotI; sex: female; dev stage: 3 months old; isolate: muscular atrophy patient; tissue type: total brain; total mRNA was oligo-dT primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lambda Z1P-Lox vector. Clone library from B. Soares, Psychiatry Dept. Columbia University, USA. Normalization method: Bento Soares, P.N.A.S. in press"

BASE COUNT 58 a 43 c 43 g 119 t 1 others

ORIGIN

Query Match 82.0%; Score 16.4; DB 21; Length 264;
Best Local Similarity 94.4%; Pred. No. 1.6e+03; Mismatches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATTTCATGAACAATGA 20
|||||
Db 103 ATTTCATGAACAATGA 86

RESULT 14
R44353/c 432 bp mRNA EST 22-MAY-1995
LOCUS YG36B02.s1 Soares infant brain INIB Homo sapiens cDNA clone
DEFINITION IMAGE:34400 3' similar to contains 11 repetitive element ;, mRNA sequence.
ACCESSION R44353
NID 9820649
VERSION R44353.1 GI:820649
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 432)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, N., Kucaba, T., Le, M., Lennon, G., Marr, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Woldmann, F., and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL MEDLINE Unpublished (1995)
COMMENT On May 9, 1995 this sequence version replaced gi:803077.

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Insert Size: 1986

*High quality sequence stops: 292 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1986 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 292.
Location/Qualifiers

FEATURES

source

```
1.3432
/organism="Homo sapiens"
/db_xref="GDB:406747"
/db_xref="taxon:9606"
/clone="IMAGE:34400"
/clone_1lb="Soares infant brain 1N1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lactid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer (5'
AAGTGAAGATTCGGCGCGCGAGGATTTTCTTTTCTTTT 3');
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lactid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT

```
98 a 70 c 88 g 174 t 2 others
```

ORIGIN

Query Match

```
Best Local Similarity 82.0%; Score 16.4; DB 22; Length 432;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 3 ATTTGATGAACAATGA 20

Db 113 ATTTGATGAACAATGA 96

RESULT 15

LOCUS

R87639

DEFINITION

```
YP89601.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:194592 3' similar to contains Alu repetitive element; mRNA
sequence.
```

ACCESSION

R87639

NTD

9946452

VERSION

R87639.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
```

AUTHORS

```
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
```

TITLE

The Washu-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

Insert Size: 361

High quality sequence stops: 267

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 361

Std Error: 0.00

FEATURES

source

```
Seq primer: Promega -21ml3
High quality sequence stop: 267.
Location/Qualifiers
1.358
/organism="Homo sapiens"
/db_xref="GDB:376362"
/db_xref="taxon:9606"
/clone="IMAGE:194592"
/clone_1lb="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
(5' AACTGGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3');
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT

```
98 a 76 c 97 g 83 t 4 others
```

ORIGIN

Query Match

```
Best Local Similarity 82.0%; Score 16.4; DB 23; Length 358;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 2 TATTCATGAACAATG 19

Db 91 TATTCATGAACAATG 108

Search completed: September 13, 1999, 15:45:56
Job time: 4044 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:44 ; Search time 539.84 Seconds

(without alignments)
111.933 Million cell updates/sec

Title: us-09-325-095-24

Perfect score: 19

Sequence: 1 AAGGATCAGACAAATGCT 19

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database: GenBank.*

1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_ba4:*
5: gb_ba5:*
6: gb_ba6:*
7: gb_ba7:*
8: gb_ba8:*
9: gb_ba9:*
10: gb_ba10:*
11: gb_ba11:*
12: gb_ba12:*
13: gb_ba13:*
14: gb_ba14:*
15: gb_ba15:*
16: gb_ba16:*
17: gb_ba17:*
18: gb_ba18:*
19: gb_ba19:*
20: gb_ba20:*
21: gb_ba21:*
22: gb_ba22:*
23: gb_ba23:*
24: gb_ba24:*
25: gb_ba25:*
26: gb_ba26:*
27: gb_ba27:*
28: gb_ba28:*
29: gb_ba29:*
30: gb_ba30:*
31: gb_ba31:*
32: gb_ba32:*
33: gb_ba33:*
34: gb_ba34:*
35: gb_ba35:*
36: gb_ba36:*
37: gb_ba37:*
38: gb_ba38:*
39: gb_ba39:*
40: gb_ba40:*
41: gb_ba41:*
42: gb_ba42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
18	94.7	3424	10_HSPH13K	229090 H.sapiens m.

2	16.4	86.3	84539	10_HSABLR3
3	16.4	86.3	4022	11_HS06616
4	16	84.2	169	9_HS169G2R
5	16	84.2	79810	42_AC005543
6	15.8	83.2	217420	1_BS08017
7	15.8	83.2	75188	7_AB005238
8	15.8	83.2	77999	7_AB005238
9	15.8	83.2	60000	8_AF049236
10	15.8	83.2	1348	8_AF093254
11	15.8	83.2	103194	8_ATAC006223
12	15.8	83.2	37730	8_HS438N21
13	15.8	83.2	44826	9_HS438N21
14	15.8	83.2	4446	9_HUMTPT11
15	15.8	83.2	4626	9_HUMTPT11
16	15.8	83.2	42896	10_HS447H23
17	15.8	83.2	114600	10_HS796F18
18	15.8	83.2	118968	10_HS8107C2
19	15.8	83.2	44665	10_HS298266
20	15.8	83.2	49999	10_HS298266
21	15.8	83.2	122322	11_AC002448
22	15.8	83.2	168056	11_AC004458
23	15.8	83.2	116019	11_AC004629
24	15.8	83.2	142278	11_AC005177
25	15.8	83.2	11046	11_HSSTRAT5X1
26	15.8	83.2	21873	11_HUAC004382
27	15.8	83.2	256	14_G36107
28	15.8	83.2	267	14_HUMC5011
29	15.8	83.2	238438	34_AC005046
30	15.8	83.2	110000	34_AC005079.4
31	15.8	83.2	223072	34_HSD3737M1
32	15.8	83.2	224064	35_AC005055
33	15.8	83.2	155093	35_AC005948
34	15.8	83.2	203234	35_AC006205
35	15.8	83.2	204544	35_AC007677
36	15.8	83.2	262929	35_AC007848
37	15.8	83.2	180388	42_AC006206
38	15.8	83.2	135021	42_AC007159
39	15.4	81.1	217000	1_AP000004
40	15.4	81.1	2050	1_STYCHEA
41	15.4	81.1	2408	3_BOVOST1
42	15.4	81.1	98822	7_ATT6K21
43	15.4	81.1	90341	8_ATAC006232
44	15.4	81.1	4337	9_MACPRB
45	15.4	81.1	119927	11_AC005062

ALIGNMENTS

RESULT	1	RESULT	1	RESULT	1
HSPH13K	HSPH13K	3424 bp	mRNA	PRI	24-AUG-1995
LOCUS	H.sapiens mRNA for phosphatidylinositol 3-kinase.				
DEFINITION	H.sapiens mRNA for phosphatidylinositol 3-kinase.				
ACCESSION	229090				
KEYWORDS	9472990				
VERSION	229090.1				
KEYWORDS	GI:472990				
SOURCE	phosphatidylinositol 3-kinase.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Euteria;				
TITLE	Primates; Catarrhini; Hominoidea; Homo.				
JOURNAL	1 (bases 1 to 1068)				
MEDLINE	Volinia, S., Hiles, I., Ormonde, E., Nizetic, D., Antonacci, R.,				
REFERENCE	Rocchi, M. and Waterfield, M.D.				
AUTHORS	Molecular cloning, cDNA sequence, and chromosomal localization of				
TITLE	the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene				
JOURNAL	Genomics 24 (3), 472-477 (1994)				
	2 (bases 1 to 3424)				
	Volinia, S.				
	Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig				
	Institute for Cancer, Research, 91 Riding House Street, London, W1P				

acute leukemias with rearrangements in the BCR gene first intron: further evidence implicating Alu sequences in the chromosome translocation
Nucleic Acids Res. 17 (19), 7631-7642 (1989)

JOURNAL MEDLINE
AUTHORS
TITLE
14 (sites)
Tatusov,R. and Lipman,D.J.
Using local similarities for pattern detection in nucleotide/protein sequences
unpublished

JOURNAL MEDLINE
AUTHORS
TITLE
15 (sites)
Groffen,J., Stephenson,J.R., Heisterkamp,N., de Klein,A., Bartem,C.R. and Grosveld,G.
Philadelphia chromosome breakpoints are clustered within a limited region, bcr, on chromosome 22

JOURNAL MEDLINE
AUTHORS
TITLE
16 (bases 37662 to 37903; 49889 to 50062; 50626 to 50719)
Shivelman,E., Lifshitz,B., Gale,R.P. and Canaan,I.E.
Fused transcript of abl and bcr genes in chronic myelogenous leukaemia
Nature 315 (6020), 550-554 (1985)

JOURNAL MEDLINE
AUTHORS
TITLE
17 (bases 43469 to 44077)
Chen,S.J., Chen,Z., d'Aurion,L., Le Coniat,M., Grausz,D. and Berger,R.
Phi/bcr- acute leukemias: implication of Alu sequences in a chromosomal translocation occurring in the new cluster region
Oncogene 4 (2), 195-202 (1989)

JOURNAL MEDLINE
AUTHORS
TITLE
18 (bases 1 to 84539)
Chisoe,S.L.
Sequence of the human abl and bcr genes
Thesis (1994) University of Oklahoma
19 (bases 1 to 84539)
Roe,B.A.
Direct Submission
Submitted (08-MAR-1994) Bruce A. Roe, Dept. of Chem. and Biochem., University of Oklahoma, 620 Parrington Oval, Rm 208, Norman, OK 73019, USA

JOURNAL MEDLINE
AUTHORS
TITLE
20 (bases 43316 to 43837)
van der Pelz,M.J., Shivji,M.K., Allen,P.B., Heisterkamp,N., Groffen,J. and Wiedemann,L.M.
Nucleotide sequence of both reciprocal translocation junction regions in a patient with Ph positive acute lymphoblastic leukaemia, with a breakpoint within the first intron of the BCR gene

JOURNAL MEDLINE
AUTHORS
TITLE
Nucleic Acids Res. 17 (1), 1-10 (1989)

JOURNAL MEDLINE
AUTHORS
TITLE
21 (bases 80055 to 82432)
Andrews,D.F., III., Tompkins,C.K., Hendrickson,S.L. and Singer,J.W.
Cloning and sequencing of the human c-abl 3' untranslated region
Oncogene 5 (3), 441-444 (1990)

JOURNAL MEDLINE
AUTHORS
TITLE
22 (bases 24740 to 25039)
de Klein,A., van Agthoven,T., Groffen,C., Heisterkamp,N., Groffen,J. and Grosveld,G.
Molecular analysis of both translocation products of a Philadelphia-positive CML patient
Nucleic Acids Res. 14 (17), 7071-7082 (1986)

JOURNAL MEDLINE
AUTHORS
TITLE
23 (sites)
Heisterkamp,N., Stam,K., Groffen,J., de Klein,A. and Grosveld,G.
Structural organization of the bcr gene and its role in the Ph' translocation
Nature 315 (6022), 758-761 (1985)

JOURNAL MEDLINE
AUTHORS
TITLE
24 (sites)
Litz,C.E., McClure,J.S., Copenhaver,C.M. and Brunning,R.D.
Duplication of small segments within the major breakpoint cluster region in chronic myelogenous leukemia
Blood 81 (6), 1567-1572 (1993)

JOURNAL MEDLINE
AUTHORS
TITLE
93200517
25 (sites)
Soekarman,D., van Denderen,J., Hoefsloot,L., Morel,M., Meuwesen,T., van Baal,J., Hagemeijer,A. and Grosveld,G.
A novel variant of the bcr-abl fusion product in Philadelphia chromosome-positive acute lymphoblastic leukemia
Leukemia 4 (6), 397-403 (1990)

JOURNAL MEDLINE
AUTHORS
TITLE
90294679
Location/Qualifiers
1..84539
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="9q34"
/gene="ABL"
join(U07561.1:29132..35962,U07562.1:1..59012.1..80445)
join(U07561.1:29132..29267,49889..50062,50626..50921,58588..58860,67954..68038,68685..68862,70893..70877,73175..73527,74828..74917,75260..75424,78731..80445)
/gene="ABL"
/note="ABL is the cellular homolog proto-oncogene of Abelson's murine leukemia virus and is associated with the t(9;22) chromosomal translocation with the BCR gene in chronic myelogenous and acute lymphoblastic leukemia; alternative splicing using alternative first exon 1b"
/citation-[8]
/citation-[10]
/citation-[11]
/codon_start=1

Query Match 86.3%; Score 16.4; DB 10; Length 84539;
Best Local Similarity 94.4%; Pred. NO. 53;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGGATCGAACAATGCC 18
Db 54910 AAGGATCGAACAATGAC 54893

RESULT 3
LOCUS HST66616 4022 bp mRNA PRI 18-SEP-1996
DEFINITION Human SWI/SNF complex 170 kDa subunit (BAF170) mRNA, complete cds.
ACCESSION U66616
NID 91549240
VERSION U66616.1 GI:1549240
KEYWORDS
SOURCE human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 4022)
Wang,W., Xue,Y., Zhou,S., Kuo,A., Cairns,B.R. and Crabtree,G.R.
Diversity and specialization of mammalian SWI/SNF complexes
Genes Dev. 10 (17), 2117-2130 (1996)

JOURNAL MEDLINE
AUTHORS
TITLE
2 (bases 1 to 4022)
Wang,W., Cote,J., Xue,Y., Zhou,S., Khavari,P.A., Biggar,S.R., Muchardt,C., Kapana,G.V., Goff,S.P., Yaniv,M., Workman,J.L. and Crabtree,G.R.
Purification and biochemical heterogeneity of the mammalian SWI-SNF complex
EMBO J. 15 (1996) In press

JOURNAL MEDLINE
AUTHORS
TITLE
3 (bases 1 to 4022)
Wang,W., Xue,Y., Zhou,S. and Crabtree,G.R.
Direct Submission
Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford University, Beckman Center B207, Stanford, CA 94305-5428, USA
Location/Qualifiers
1..4022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="Jurkat T cells"

gene 23..3664
/gene="BAF170"
CDS 23..3664
/gene="BAF170"
/note="similar to human BAF155 and yeast SWI3; contains a region similar to the DNA binding domain of myb and a predicted leucine zipper domain; the C-terminus of the protein is highly proline-rich"
/codon_start=1
/product="SWI/SNF complex 170 kDa subunit"
/protein_id="AAC50694.1"
/db_xref="pid:g1549241"
/db_xref="gi:1549241"
/translation="MAVRRKKGDPNRYTEADPTQEDNRLMLCKNKYITQAEPP
TKNSLSLVOLQOEVEVGRKVNAPLTKLPICQFLDKAGSGSCIIIAAYKRS
DGMRYDYFQNSRNDNEMETIEKSLVQNNKSRPFIICPEIEPIELGKNDIT
KRGHGVEDAPKNNASHVYVPGNLEEEVNRVYKRDVOLLHMGYDSDYDTWPA
SEIASVEDAPTPKPRKRYHAKWIIDTDFNMENMEDYEDVDNKPVRKRSKAT
LIDVANSPPSDRDKGCKNKKRSPSPSPPEYEECKGKSPITPKSKGRHEE
EEDDLTKDESPVPNVEEVLPTVNRKNSAPVGVGTITDIEDSDSMTTG
KDEDSTNGKEQTKNPDLHEDNTEQTHIIPSYAMFYNSVHALERALPEPF
NGKNSKTEIETLAYENMIDYRLAPQELTSGRLAGVCAISVHAFLEQWG
LNYQVDAERPTPMGPPTSHRHVADTPSGGLVLPQPTPOOTSQOOLNLPDQK
EAPDMONGELTDMYTKNAPSKSKAAATREMTDETLILLALEYKNDMNV
EYVGSRTQDECTLHLRLPIEDPILEDSSAGPLAQPIPSQSGNPMSTYAPLAS
VDPYVAAAKSALBEESKMEVPTALVEHVRVEEAAVGTGADPAFGLESIGI
AGTSDPERIESGNDARVGAQATDEKEREPEEGGALIEEKERTSAPKDE
EKGKEDSEKESKSDGPIDVPEKEKEKEGQEVLEKVESEGRKTKVERDIDG
NSTAAGDAAALAAVAKHLAAVEERIKSLVALVETOMKLEILRIHELEETMD
RERALEYQROOLADROAFHMOLEKYPMBARQOHQOHOQOOPPALPGSOP
PERTGAPRAGVAGIAPASVYAPAGSAPGSGPSQIQAGSTRPOOQAPGA
PPGAVPPGVPVPPGPHGSPPTNQTPTPSMAGAVGSGHPPAGNAPGLPGMPP
PPPPAPSIIPFGSLADISINLPAPNLPSPSPVPPPTLPPLNLPVMANPLHNL
PATTPMPSLIPGPGISAAQSPAIVAAVOGLNLPASAPLDPPTLPDPTAPSPG
TVPVPPPPQ"

BASE COUNT 1084 a 1153 c 1051 g 734 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 11; Length 4022;
Best Local Similarity 94.4%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGGATCAGAACATGCT 19
|||||

Db 528 AGGATAAGAACATGCT 545

RESULT 4
H516962R/c 169 bp DNA PRI 19-OCT-1995
DEFINITION H.sapiens Cpg island DNA genomic MseI fragment, clone 169g2,
ACCESSION reverse read cpg169g2.rta.
NID 259694
VERSION 259694.1 GI:1031607
KEYWORDS Cpg island; genomic MseI fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 169)
Doddsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.
Direct Submision
Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RO, England. E-mail contact: humquerry@sanger.ac.uk
2 (bases 1 to 169)
Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
Purification of Cpg islands using a methylation DNA binding column
Nat. Genet. 6 (3), 236-244 (1994)
JOURNAL 94282070
MEDLINE
COMMENT Vector: pGEM-5zf(-)
Clones are available from the UK MRC Human Genome Mapping Project

Resource Centre, Hinxton, Cambridgeshire CB10 1RO, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biohelp@hgmp.mrc.ac.uk.
Location/Qualifiers
1..169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/dev_stage="adult"
/tissue_type="blood"
/clone_1lb="CGI-1"
/clone="169g2"

BASE COUNT 41 a 46 c 32 g 47 t 3 others
ORIGIN

Query Match 84.2%; Score 16; DB 9; Length 169;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGATCAGAACATG 16
|||||

Db 108 AAGGATCAGAACATG 93

RESULT 5
AC005543 79810 bp DNA PRI 26-FEB-1999
LOCUS Homo sapiens chromosome 4 clone B320D6 map 4q25, complete sequence.
DEFINITION AC005543
ACCESSION 94309859
NID AC005543.2 GI:4309859
KEYWORDS HTG.
VERSION human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 79810)
Stone,N.E., Schmutz,J.J., Shang,J., Cox,D.R. and Myers,R.M.
Direct Submision
Submitted (26-AUG-1998) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
3 (bases 1 to 79810)
Stone,N.E., Schmutz,J.J., Shang,J., Cox,D.R. and Myers,R.M.
Direct Submision
Submitted (26-FEB-1999) Department of Genetics, Stanford Human
Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA
On Mar 1, 1999 this sequence version replaced g1:4225912.
Quality: Phrap Quality >=40 98.9% of Sequence;
Estimated Total Number of Errors is 0.8.
Location/Qualifiers
1..79810
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="B320D6"
/chromosome="4"
/map="4q25"
/clone_1lb="Caltech/Research Genetics Whole-Genome BAC
Library"

BASE COUNT 24697 a 15875 c 15934 g 23304 t
ORIGIN

Query Match 84.2%; Score 16; DB 42; Length 79810;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGATCAGAACATG 16
|||||

Db 9241 AAGGATCAGACATG 9256

RESULT 6
BSUB0017
LOCUS BSUB0017 217420 bp DNA BCT 26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 17 of 21): from 3197001 to 3414420
ACCESSION 299120 AL009126
MID 92635613
VERSION 299120.1 GI:2635613

KEYWORDS
SOURCE
ORGANISM

Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;

REFERENCE
AUTHORS

1 (bases 1 to 217420)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S., Boriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Conerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fume, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golding, E.J., Grand, A., Guisepi, G., Guy, B.J., Haga, K., Haeck, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaer, Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningsstein, G., Krogh, S., Kunano, M., Kurita, K., Lapius, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ojawa, K., Ogiwara, A., Oudega, B., Park, S.H., Patro, V., Pohl, T.M., Portelle, D., Porciliak, S., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scottone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serrito, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpetra, P., Tognoni, A., Josato, V., Uchiyama, S., Vandebol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and Danchin, A.

TITLE The complete genome sequence of the gram-positive bacterium

JOURNAL Nature 390 (6657), 249-256 (1997)

MEDLINE 98044033

2 (bases 1 to 217420)

AUTHORS Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.

TITLE Direct Submission

Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Reproduction de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

FEATURES
source location/Qualifiers

1..217420

/organism="Bacillus subtilis"

/strain="168"

/db_xref="taxon:1423"

complement(3..1277)

/gene="yule"

complement(3..1277)

/gene="yule"

/function="unknown"

/note="similar to L-thiamine isomerase"

/codon_start=1

/transl_table=1

/protein_id="CAB15107.1"

/db_xref="PID:el184196"
/db_xref="PID:g2635614"
/db_xref="GI:2635614"
/db_xref="SPTRMBL:005264"
/translation="MTIKANYDSAKQYERKMGIDVEERLQLEQVPIHICWOGDIE
GFEVAKELSGSIDVTGNYPKQTPPELRDLEKALSLIPGKRVLYALVETNE
AVRDELKPOHFENYVWAKNLGLDGNFLSHEKAAGLTLSPDDIPREYH
CIACRRIGETFGKEGLTPTCLNTNHPGVDITSDRTPPKRLKESIDRFSEISQ
HNDSIESKLEGLSESIVVGSHEFYALVATNKLCLDGTGHPHETYSNKTSSKL
LYTDKLAHVSAPYRMDSDHVVYLDLELALTEIVNHALERVAIGLDPEFDSINIV
AAWTIGRNMKALYALPLPNVYIKQLOEGERTERLALMEFKTYPFGAIDWDSYE
OMGYPVEAMLVDIKREYQVLLKRNASSPIV"
complement(1303..1617)
/gene="yuld"
complement(1303..1617)
/function="unknown"
/note="similar to hypothetical proteins"

CDS

gene

/codon_start=1
/transl_table=1
/protein_id="CAB15108.1"
/db_xref="PID:el184197"
/db_xref="PID:g2635615"
/db_xref="GI:2635615"
/db_xref="SPTRMBL:005263"
/translation="MKRKASIMCVHODKYEYKORHDDIPENAEALKAHGHYSIF
LDERTGLFAYLEDEDEKRNKADIEVCORWKSNAPLKNTSDSPVALDKVEY
LD"
complement(1635..3092)
/gene="yulc"
complement(1635..3092)
/function="unknown"
/note="similar to rhamnulokinase"

CDS

gene

/codon_start=1
/transl_table=1
/protein_id="CAB15109.1"
/db_xref="PID:el184198"
/db_xref="PID:g2635616"
/db_xref="GI:2635616"
/db_xref="SPTRMBL:005262"
/translation="MTYTAIDVGSGRINWGLNECKDIOEIHPRANGESORHGC
LMDIDLKLOGLQKVLGKRGCVGIDTMAVDYVLDDEKDRLEASYSRDT
TGKIDLEHLTSKALITOKTIGLOPENTITOLFEEDRELUKTDKIMIPDYGICL
DKAVTEITVSTQTLNVSTGNDELLEAVSYLEQFAPLEPGCEGLKRWEP
DYDLPAKVYVAVADYASAVIAPGVNDGMAVSSGTSWLGVENTPTITDLND
NYTERGANNYIRFLKNIGMNVIOEVKQLOADYSFOQLAEBAKTEPQOFTNLD
KRLNPNMKIEIOHGYOTROKIPRAGELACCIYATNLAITYAIKKELETIERPI
EORHITGGARNDELNLTADMSKATYAPPIEATATGNLIMQIAKVKYDIKEARQ
VVRNSPFIKVFIPKIDRSTTIOSTQVTLKALSK"
complement(3097..3873)
/gene="yulb"
complement(3097..3873)
/function="unknown"
/note="similar to transcriptional regulator (Deor family)"

CDS

gene

/codon_start=1
/transl_table=1
/protein_id="CAB15110.1"
/db_xref="PID:el184199"
/db_xref="PID:g2635617"
/db_xref="GI:2635617"
/db_xref="SWISS-PROT:005261"
/translation="MLVAERQKIVEIVNRRSSIRVSLDIFSVEETIRRDLEKLE
KEHLSRHGGAAGVIOOKESIHSESEINIVLEKKAIAHAARVYVSGRIIDAST
TAYMAKIIDIELTETVTSNKAIEISNENIVSITGILKESISFVGPIAEREL
ETVHNKTLFSCGFVHNKSDSNENQALIKKMIERSDQTLIADSSKRGNEFSH
IASIQDVSRLITDSGLDPASVKALEDKRVAVTAVPLSKRG"
complement(3930..5999)
/gene="yuxg"
complement(3930..5999)
/function="unknown"

CDS

gene

```

/note="alternate gene name: yula; similar to
scb1ol-6-phosphate 2-dehydrogenase
/codon_start=1
/transl_table=11
/protein_id="CAB15111.1"
/db_xref="PID:el184200"
/db_xref="GI:2635618"
/db_xref="GI:2635618"
/db_xref="SWISS-PROT:P40747"
/translation="MYKHIWDSERAAOLPRGVEELVYRSNLIGSDRTVCNMGNGNTSM
KTEKDFRGREIEVMVWGSGDLATMKANFSGKLDDIRPLIKDQMPDEMVDYL
SHOMDSKRPSPRIETLHAPLYKHVDHPDAIISTICANGKQIADIYGRBVM
VYVVRGRTKPSKRIAGVANNPRAEVLMEKGIYWGTSSECTYKTIISTIQBAEQY
INDRINQHEVFGKRTOPLEDRKQILGIMPEVINGA/SEEKMLISTDDHDLVEF
VNSVQAPALSQIGACPDHLYHKRVPILDMNPETQVHKLADLKRSGEVEFTSEYQ
AVTRNOQDGDOLFESAPRVLIPGIGVNTGVSYSMSVSGALYRRAIAVKAGATL
GOFVSLHNESEYHVEWPLELKLTLAPAEAESRKVALITGAGAGISAACRRFAE
GGHVIYADNINIGAKOIKAGEINDAVGKGRAMVKNMVTKEPDVQSAFEBAALAYGID
IYVNNAGLATSSPDETSKERNLANNVIGTGYELARARAFQMKHONGSGMVPFGS
KNSVYAGKASASVKALETHLARCIAEGSGFIRVSVLPDVALQCSALWGSWR
ERRAAVGIPEPOLSEHYRRKRLALVNIY PEDIAIAEPASSKAEKTTGCMITVDGG
VPAAFTR"
/complement(6109..8125)
/gene="tlpB"
/complement(6109..6124)
/gene="tlpB"
/complement(6137..8125)
/gene="tlpB"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein"
/protein_id="CAB15112.1"
/db_xref="PID:el184201"
/db_xref="PID:92635619"
/db_xref="GI:2635619"
/db_xref="GI:2635619"
/db_xref="SWISS-PROT:P39217"
/translation="MGKFTQMTKOPISITPLIAFLAVLILPVGLVAFYSOSANMAL
DELLSSAKGVNELNSTQNKLEDRKAIIDYSEYVDLILKRRKTIKKEKFOYT
TINDVGYIYANASEDKLTRYPDGVPKGFDPGRDMYKQAVAKQOAFSEPTDEA
TGDIIVYTIKQLKDSGVIALDLNDELVTASRKIRIGRGAFTTGKMKYIAHPTI
KPGTGSQDMTQNVYSGKEGSEYEPGEKEMKAPTNLTGKMTAGTEFVSELDOAS
SPVLNVAIILCVSIYIGILITLYIRATTKPKLYSASAKISSGIDLEVIDISKNS
EFGOLGSEFNEMASLRVYIGVYOTSEVNASSSELTASAOATSKATRHILALAEF
SPGNEAOSKLETSNHLSONMGDISKVAQASTITKKSIOSEASGSEKRYEHTYG
OKRTIDQYQKAQAVVKGLETSQDITSLNLTNGADTNTLLALALAEARAGETG
RGSVVAEVRKLAQASADSAKEIEGLIDEIVAREISLSMEQSVNHEKELQITDQ
TAESEFOIYEMTQISGEQNLNATVEQLSAGSOEVSAYVEDISAVAKESGAGIODIA
ASAEQOLASMEETISSAETLANNAEELADITKKFKIES"
/complement(8204..8228)
/gene="mcpA"
/complement(8204..10224)
/gene="mcpA"
/complement(8239..10224)
/gene="mcpA"
/function="chemotaxis deficiency toward glucose and
alpha-methyl-glucoside"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein"

```

Query Match 83.2%; Score 15.8; DB 1; Length 217420;
 Best Local Similarity 89.5%; Pred. No. 1.3e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGATCAGACATGCT 19
 ||||| ||| ||||| |||||
 Db 133918 AAGGACAGCAATGCT 133936

RESULT 7
 LOCUS AB005238 75188 bp DNA PLN 05-FEB-1999
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MKP11,

```

ACCESSION complete sequence.
AB005238
NID 9264310
VERSION AB005238.1 GI:2264310
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1
clone:MKP11.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (sites)
REFERENCE
AUTHORS Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M.,
Miyajima,N. and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. I.
Sequence features of the 1.6 Mb regions covered by twenty
physically assigned P1 clones
JOURNAL DNA Res. 4 (3), 215-230 (1997)
97471969
2 (bases 1 to 75188)
REFERENCE
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1997) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
(E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
Fax:+81-438-52-3934)
FEATURES
source
1..75188
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MKP11"
/location="lib:Mitsui P1"
BASE COUNT 24812 a 12842 c 13514 g 24020 t
ORIGIN

```

Query Match 83.2%; Score 15.8; DB 7; Length 75188;
 Best Local Similarity 89.5%; Pred. No. 1.2e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGATCAGACATGCT 19
 ||||| ||||| ||||| |||||
 Db 33173 AAGGACAGCAATGCT 33191

RESULT 8
 AB005246/C
 LOCUS AB005246 77999 bp DNA PLN 05-FEB-1999
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MUP24,
 complete sequence.
 ACCESSION AB005246
 NID 9264318
 VERSION AB005246.1 GI:2264318
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1
 clone:MUP24.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsids.
 1 (sites)
 REFERENCE
 AUTHORS Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M.,
 Miyajima,N. and Tabata,S.
 TITLE Structural analysis of Arabidopsis thaliana chromosome 5. I.
 Sequence features of the 1.6 Mb regions covered by twenty
 physically assigned P1 clones
 JOURNAL DNA Res. 4 (3), 215-230 (1997)
 97471969

REFERENCE 2 (bases 1 to 77999)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1997) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2: 1532-3, Yama, Kisarazu, Chiba 292, Japan
(E-mail:nakamura@kazusa.or.jp, Tel:+81-438-52-3935,
Fax:+81-438-52-3934)

FEATURES
source
1. .77999
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MCP24"
/clone_id="Mitsui P1"
BASE COUNT 24666 a 13395 c 14042 g 25896 t
ORIGIN

Query Match 83.2%, Score 15.8; DB 7; Length 77999;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGATCAGACAAATGCT 19
|||||
DB 12726 AAGGATCAGAACAAATTTCT 12708

RESULT 9
AF049236/c
LOCUS AF049236 60000 bp DNA PLN 22-APR-1998
DEFINITION Arabidopsis thaliana putative transmembrane protein gip (Atg1),
putative nuclear DNA-binding protein G2P (Atg2), Em1 protein
(ATEM1), putative chlorophyll synthetase (Atg4), putative
transmembrane protein G5P (Atg5), putative acyl-coA dehydrogenase
(Atg6), and calcium dependent protein kinase genes, complete cds;
and unknown genes.

ACCESSION AF049236
NID AF049236
VERSION g3068702
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 16488 to 21370)
Gaubier,P., Wu,H.-J., Laudie,M., Delseny,M. and Grellet,F.
A chlorophyll synthetase gene from Arabidopsis thaliana
Mol. Gen. Genet. 249 (1), 58-64 (1995)
96140448
2 (bases 1 to 60000)
Gaubier,P., Wu,H.-J., Berger,C., Cooke,R., Delseny,M. and
Grellet,F.
Sequence of a 60kb contig on chromosome 3, the ATEM1 locus
Unpublished
3 (bases 1 to 60000)
Grellet,F.
Direct Submission
Submitted (19-FEB-1998) LDBMP, Universite de Perpignan, 52 Avenue
de Villeneuve, Perpignan 66860, France
Location/Qualifiers
1. .60000
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/complement(join(922..1362,1541..>1786))
/product="hypothetical protein"
/complement(join(922..1362,1541..1786))
/note="hypothetical protein; ORF1: similar to sequence in
Genbank Accession Number AC002330"
/codon_start=1.

FEATURES
source
1. .60000
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/complement(join(922..1362,1541..>1786))
/product="hypothetical protein"
/complement(join(922..1362,1541..1786))
/note="hypothetical protein; ORF1: similar to sequence in
Genbank Accession Number AC002330"
/codon_start=1.

/product="unknown"
/protein_id="AAC14403.1"
/db_xref="PID:g3068703"
/db_xref="GI:3068703"
/translation="MILYLVVEEMDINRSRLNENAPLSTKRHYOQKLPQGFPA
SYPCNSLACGMPKNSHLLKNDAESEPRLKPGRAKEPHQNDQAVYLSQTAV
DMVLPKDSYSESDRSVGGDSNGILTESETERSDSPSTIIERLPDAALAAV
TSAQSRKRLSTLSETVRSQCFSPKAGLHVLLPWSKTRHRCGVKNMSPSSVH
LQPFIRKDN"
complement(join(<2709..2935,3332..3641,3930..4421,
4329..>6376))
/product="hypothetical protein"
complement(join(2709..2933,3332..3641,3930..4421,
4329..6376))
/note="hypothetical protein; ORF2: similar to sequence in
Genbank Accession Number AC002330"
/codon_start=1
/product="unknown"
/protein_id="AAC14404.1"
/db_xref="PID:g3068704"
/db_xref="GI:3068704"
/translation="MCHNLEFTMRSLKLAEGKGTQYVALNPSPAPPPGNSGTG
GGGGSGGTGGVGDKLQHLSDHLRYNSVRSKSRTPPTOPNAVSPFELPGL
PYDLELPDIDPCLFYDIYERKMAQVYRIENCSQEPKSGAYLEQALIFRISPKLF
RRSLRSRQAHVDYNAKAVLASWLSGGBELIGTSMDCGRNLECPKATVSGTTP
ESVYDPCVSGASRSEMNDECTSOEVYDMSFCIGDEVRVRYKLASRPEFA
MLYGFREMKRATINFTONGISVEGMRAELFEFTNLDNFPVAVLEKLARNPCC
DEKSGCDSHLALVNSLDLDEAMLLIEGYLAVALVLAACLOVLRLLPSMHPNYI
KIPCSAEGRERLAIGHASFTLYPFLSIAMEDDMKSNTRYMLLERVECAVDSMEO
LAYHOLGVVLEREYEDQADQNRMAAYEAHLYLVLVARTPKRDRYSAYKIINL
ISDRKATGMHQRSLYCSGERKLLDIDITEPDTITFPKFAVALVEHQGAL
AELNRKILGFASPOCLERAMVISTGMEDYEGALDIALLTLEENFMKPKHGDH
VELRLPAAQMSQDCWMOLOYDRWSVDIGSLAVVHMLANDPGKILFRKSLLL
RLNCKAKAMSLRLAARNSKSEHRLVYEGMIIYDTGHEEALAKAESISIQSFPA
FLKAVYALSTLDPDSSNVVIOLGALICPSGLKGAALNIGSVYDCELDLIA
ADCYTNLTIKRPAHOGIARVYLKORRAADPEMKLIEKANNASAYEKRESCD
REKASDLCIATQIDPRTPIRYRRAAVLMDHRESAIDELSLAISFKPDQLHHR
AAFYDSMGEASAIRKDEEALCIDPGADVLELYHKAREPDOK"
complement(join(9320..9605,9694..9840,9918..10174,
10350..10662))
/product="hypothetical protein"
complement(join(9408..9605,9694..9840,9918..10174,
10350..10537))
/note="hypothetical protein; ORF3: similar to Arabidopsis
thaliana clone 240M377 sequence found in dbEST Accession
Number N65910"
/codon_start=1
/product="unknown"
/protein_id="AAC14405.1"
/db_xref="PID:g3068705"
/db_xref="GI:3068705"
/translation="MMNSTEESMEYRPGMLYORRDPASSDHRKLPDPSANAF
AOTRITVSGSHSHDLHISAHAFPGVKAAYKQGLSEAKLTIRGVRDRAOL
OAGVYKASLYVVEPTNKRVDOQPPVYTKEMKALAAVNAATGEVDKSDRYALE
VAVNGTQVAVREFDMAEILMRQLLITIEAGDAKAVRKVEAYIDNQVAVDL
KARCSNPFVDSKAAAVSTEMESGNGVSLNPPASPANVTQDMERD"
complement(12025..13047)
/gene="Atg1"
complement(<12272..>13042)
/gene="Atg1"
/product="putative transmembrane protein gip"
complement(12272..13042)
/gene="Atg1"
/note="similar to Bradyrhizobium japonicum CycD: SwissProt
Accession Number P45401"
/codon_start=1
/product="putative transmembrane protein gip"
/protein_id="AAC14406.1"
/db_xref="PID:g3068706"
/db_xref="GI:3068706"
/translation="MAARLLFRSSQILRSIORNPQISSSESPDPIFSLTASP
PSRSLSTFLRSLSIARGPTRRKIDIGAKARQMDNRMLTAYLTITSCAGFVYIL
NODQDLVLTLPDAMEKFAENPTKPKRILGVLLEGVAQPAQASQEMEFVITDILT

Query Match 83.2%; Score 15.8; DB 8; Length 1348;
 Best Local Similarity 89.5%; Pred. No. 74;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGATCGACAGACAGCCCT 19
 ||||||| |||||
 DB 256 AAGGATCGACAGATTCCT 274

RESULT 11
 ATAC006223 103194 bp DNA PLN 20-FEB-1999
 LOCUS Arabidopsis thaliana chromosome II BAC F22D22 genomic sequence,
 DEFINITION complete sequence.
 ACCESSION AC006223
 NID 94263694
 VERSION AC006223.2 GI:4263694
 KEYWORDS HTG.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euophyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsids.
 1 (bases 1 to 103194)
 Lin.X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., Vanaken,S.E.,
 Barnstead,M.E., Mason,T.M., Bowman,C.L., Rongning,C.M., Benito,M.,
 Carrere,A.J., Greasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,
 Fraser,C.M. and Venter,J.C.
 Arabidopsis thaliana chromosome II BAC F22D22 genomic sequence
 Unpublished
 2 (bases 1 to 103194)
 Lin.X. and Kaul,S.
 Direct Submission
 Submitted (15-DEC-1998) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA, xlinetlgr.org
 3 (bases 1 to 103194)
 Lin.X.
 Direct Submission
 Submitted (20-FEB-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Feb 24, 1999 this sequence version replaced gi:4176358.
 Address all correspondence to:
 Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr.
 Rockville, MD 20850, USA
 e-mail: xlinetlgr.org
 BAC clone F22D22 is from Arabidopsis chromosome II and is near the
 molecular marker Tm5.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.
 Genes were identified by a combination of three methods: Gene
 prediction programs including GRAFT (available by anonymous ftp
 from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
 Washington), Genscan (Chris Burge,
 http://genomic.stanford.edu/~chris/GENSCANW.html), and NeuplatGene
 (http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (http://www.tigr.org/tdb/at/ac.html).
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
 http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
 identified by RepeatMasker (Arian Smit,
 http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of

FEATURES

genomic sequence that are not annotated as genes but have predicted
 exons by GRAFT are annotated as misc features.

source

1..103194
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="II"
 /map="Tm5"
 /clone="F22D22"
 1..1596
 /note="overlap with BAC clone T32F6
 (AC005700:98829..100424)."
 join(<80..133,321..3476,3704..3808,3904..4059,4278..4376,
 4557..>4796)
 /gene="F22D22.1"
 <80..>4796
 /gene="F22D22.1"
 /note="13256-24"
 join(80..133,321..3476,3704..3808,3904..4059,4278..4376,
 4557..4796)
 /gene="F22D22.1"
 /codon_start=1
 /product="putative myosin II heavy chain"
 /protein_id="A015381.1"
 /db_xref="pid:94263695"
 /db_xref="gi:4263695"
 /translation="MEEAQVTSSEVPYKGDROCRKCLDLQAVNGEVPKEKEEE
 DGEFFKVEKAFDAKDAKADHVEVEKQETIEKESGSOEELHEGQEKLELEL
 EVRGEELKRYSENPDLDELAKRELEKREKGGDELVQKKQKRIIVEEESHSS
 QKSLDELQSDAKDKELTEYKRFADLGLTELESRRKKLIELEGKRSAREAKPE
 ELHKQSAHDSQKALEFSELKSTESKAKEMEKASLQOEIKELNKSSEKVE
 EALKSSAGELAAVQDELALSRLTEQVYSSTELIDELTEDEKRLTANAKLA
 ELSVLDQDQKGLQAKLSEQKINSKLAELKELLESLSKQDEKRLTANAKLA
 EYKKEBALBNVAEVTGNVAETVEVCEKLEKLEKLESDENFETALSOANNSSE
 EOKLKSLELSENGSAAATQKRLFEQVYRSSQAAFEKSQLKELETFYTAQ
 KNALEQDLNLQKSSDAERKELKSSSELDQRAIEVAEERKQATQNOEYKQA
 SELESLTQSSSRNSLEEDLNLQKAEHEDRANTTQRTLEGLQSSQSHED
 AEGRLKDELILQTERKRIQELQEVNSLEKKGTEADSKYLQVAVELSTLAFQ
 VSSSLLEALNINATENKELTENLNAVSEKKEKATVDEYSVKISEENLSELSRNE
 LVNTOGKLESTENDELKAAGLOSEYMEKLSAEELDEKGRIDEATTKRMSELHQ
 SLSDGHEHRLKAMEEPTSRDSEASLTERKLDGRTKSYEEOLEAFASGSSILKEK
 LEQIGRIAAAEVYKIKQEPDQAKESLOSSSESLAEFNNQKIKICQELBELIG
 SGSEVKEKRLKLELAIERFQKETSDDLVEKLTENQUBETIKLHAHSGVADIR
 KVELDALSKLKNLESTIEELGAKQGLKESGDLAEVNLKINLELVANHSANLQ
 KLSALEAEKQDANLELASKTIIDELTQKLSQEKLSQIEKLAVAAEKSVLESHF
 ELETKELESEVKAOLKENYENNAATASVKAELTSKLOEHEHLAGEDVLEQVLOKRE
 LOAAGSIDDEOKRLKRLRLSINLNMOKIIRKSSOMKRLRLETEAMDGVKSRDI
 DISPSPTKRKSKKPEASLSSSSSSGAVTTPQYATSHLMTVIVTVALISVIIG
 IILGRKY"

misc_feature

mRNA

gene

CDS

misc_feature
 5226..5915
 /note="exon predicted by xgrail, quality excellent"
 complement(5283..5553)
 /note="exon predicted by xgrail, quality
 excellent_shadowexon"
 complement(5638..5906)
 /note="exon predicted by xgrail, quality
 excellent_shadowexon"
 5067..6180
 /note="exon predicted by xgrail, quality excellent"
 complement(6356..6536)
 /note="exon predicted by xgrail, quality good_shadowexon"
 6423..6587
 /note="exon predicted by xgrail, quality excellent"
 join(<7175..7269,7839..8808,8890..>9075)
 /gene="F22D22.2"
 <7175..>9075
 /gene="F22D22.2"
 /note="predicted by genefinder"
 join(7175..7269,7839..8808,8890..9075)
 /gene="F22D22.2"
 /note="hypothetical protein"
 /codon_start=1

Category	Feature	Value
misc_feature	misc_feature	Protein_id="AAD15382.1"
	misc_feature	/db_xref="pid:g4263696"
misc_feature	misc_feature	/translation="MYNPQAVSTVVFISFIPPERKFLSGSHLCTLPAAAG SASPSSENLSRAKKAITQSPSPALLKQKIDWOSKQDVLEALRLDEARNGVLSO YHYNVLVCSLAEATESSPNPGLSRGDFIDFQOMVDKRPPEALFTNARLAVAD DPEAFDMYKQKMAFGIQPRLRSGYALFEFCRGADKAYEDVAHVEEVEEPE LAAALKVSHDTRKADVGYTLQRLDRLVROVSKSTEDIMEKSEVATGVYKMY KIRRAVVGSGGNGHOGVLTGKMYNRTREREMDNGYCKCKEKLIVIDINPTEFE ASLIRLACEYEVKANFNOFQELTHERPPDAVIDGANNGLVNRSSFFQVVLTLSS SNALKNHCELTTLISNLTPLA"
	misc_feature	9352..9472
misc_feature	misc_feature	/note="exon predicted by xgrall, quality good"
	misc_feature	complement(9510..9634)
misc_feature	misc_feature	/note="exon predicted by xgrall, quality marginal,shadowexon"
	misc_feature	9547..9618
misc_feature	misc_feature	/note="exon predicted by xgrall, quality excellent"
	misc_feature	9704..9771
misc_feature	misc_feature	/note="exon predicted by xgrall, quality excellent"
	misc_feature	complement(<10058..>10465)
misc_feature	misc_feature	/gene="F22D22.3"
	misc_feature	complement(<10058..>10465)
misc_feature	misc_feature	/gene="F22D22.3"
	misc_feature	complement(10058..10465)
misc_feature	misc_feature	/gene="F22D22.3"
	misc_feature	complement(10058..10465)
misc_feature	misc_feature	/codon_start=1
	misc_feature	/product="putative ribosomal protein L27"
misc_feature	misc_feature	/protein_id="AAD15383.1"
	misc_feature	/db_xref="pid:g4263697"
misc_feature	misc_feature	/db_xref="gi:4263697"
	misc_feature	/translation="MYKCAKPGKAVILLGRTYGKAAVYKSDGQTVERRKHCLVA GIKRTPSKVRKDSAKTKAKSRVKCFEYVINGQHPMTXYTLIDLKKNVSADASS KDKVYALKEAKAEERERKTRNRPFKLR"
misc_feature	misc_feature	complement(1010110757..1090111017..1115511303..11383)
	misc_feature	/gene="F22D22.4"
misc_feature	misc_feature	complement(10757..11383)
	misc_feature	/gene="F22D22.4"
misc_feature	misc_feature	complement(10757..11383)
	misc_feature	/gene="F22D22.4"
misc_feature	misc_feature	complement(1010110856..1090111017..1115511303..11333)
	misc_feature	/gene="F22D22.4"
misc_feature	misc_feature	/note="unknown protein"
	misc_feature	/codon_start=1
misc_feature	misc_feature	/protein_id="AAD15384.1"
	misc_feature	/db_xref="pid:g4263698"
misc_feature	misc_feature	/db_xref="gi:4263698"
	misc_feature	/translation="MSQISQNSGCAFPYPPVSTGPVAPPLGYPPTNDTSHATVATV ETKSGDGLKGLCAAMCCCVADAC"
misc_feature	misc_feature	12175..12251
	misc_feature	/rpt_family="TAAA"
misc_feature	misc_feature	complement(1010112340..1241912550..1267612826..>12856)
	misc_feature	/gene="F22D22.5"
misc_feature	misc_feature	complement(12340..>12856)
	misc_feature	/gene="F22D22.5"
misc_feature	misc_feature	complement(1010112371..1241912550..1267612826..12856)
	misc_feature	/gene="F22D22.5"
misc_feature	misc_feature	/note="unknown protein"
	misc_feature	/codon_start=1
misc_feature	misc_feature	/protein_id="AAD15385.1"
	misc_feature	/db_xref="pid:g4263699"
misc_feature	misc_feature	/db_xref="gi:4263699"
	misc_feature	/translation="MSQISQNSGCAFPYPPVSTGPVAPPLGYPPTNDTSHATVATV EADGFLKGLATMGLACCVADAC"
misc_feature	misc_feature	complement(13267..13307)
	misc_feature	/note="exon predicted by xgrall, quality marginal"
misc_feature	misc_feature	complement(13411..13466)
	misc_feature	/rpt_family="TA"
misc_feature	misc_feature	13633..13676
	misc_feature	83.2%; Score 15.8; DB 8; Length 103194;
misc_feature	misc_feature	Best Local Similarity 89.54; Pred. No. 1.2e+02;
	misc_feature	Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 92783 AGAATCATACAATGCCT 92801

RESULT	12
YSC18543/c	
LOCUS	YSC18543
DEFINITION	37730 bp , DNA
ACCESSION	Saccharomyces cerevisiae chromosome XII cosmid 8543.
VERSION	U20618.1
KEYWORDS	GI:2258165
SOURCE	
ORGANISM	
REFERENCE	
1	baker's yeast strain-S288C (AB972) .
2	Saccharomyces cerevisiae
3	Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
4	Chasee J et al 1997

REFERENCE	AUTHORS
(Pages 1 to 37730)	
	Johnston, M., Hillier, L., Riles, L., Albermann, K., Andre, B., Anstoye, W., Benes, V., Bruckner, M., Dellus, H., Dubois, E., Duhamel, A., Entlan, K.-D., Floeth, M., Goffeau, A., Hebling, U., Heumann, K., Heuss-Neltzelt, D., Hilbert, H., Hilger, F., Kleine, K., Koller, P., Louis, E. J., Messenguy, F., Mewes, H. W., Miosga, T., Mostl, D., Muller-Auer, S., Nentwich, U., Oberbauer, B., Plavandt, E., Pohl, T. M., Portetelle, D., Purnelle, B., Reichmann, S., Rieger, M., Rink, M., Rose, M., Schafte, M., Scherens, B., Scholler, P., Schwager, C., Schwarz, S., Underwood, A. P., Uristiazu, L. A., Vandenbol, M., Verhaesselt, P., Vierendeels, F., Voel, M., Volckaert, G., Voss, H., Wambutt, R., Wedler, E., Wedler, H., Zimmermann, F. K., Zollner, A., Hantl, J. and Hohelsel, J. D.
TITLE	The nucleotide sequence of <i>Saccharomyces cerevisiae</i> chromosome XIII
JOURNAL	Nature 387 (6632 Suppl.), 87-90 (1997)

TITLE	The sequence of <i>S. cerevisiae</i> cosmid 85443
JOURNAL	Unpublished (1994)
REFERENCE	3 (bases 1 to 37750)
AUTHORS	Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (03-FEB-1995)
REFERENCE	4 (bases 1 to 37750)
AUTHORS	Cherry, J.M.
TITLE	Direct Submission
JOURNAL	Submitted (15-JUL-1997) <i>Saccharomyces</i> Gen
COMMENT	On Jul 16, 1997 this sequence version replaced

JOURNAL COMMENT
Submitted (16-Jul-1997) Saccharomyces genome Database
On Jul 16, 1997 this sequence version replaced gi:562124.
Submitted by: Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA
e-mail: mjsequenceur.wustl.edu

NEIGHBORING COSMID INFORMATION:

This sequence includes nucleotides 1-37729 of cosmid 8543. It overlaps with the lambda clone on the left (2142) by 2300 nucleotides, and with the cosmid on the right (8300) by 200 nucleotides.

FEATURES	Location/Qualifiers
source	1. .37730

```

/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/map="XII"
/strain="S288C (AB972)"
/complement(<1..2007)
/gene="CDC25"
/complement(<1..2007)
/gene="CDC25"
/note="YLR310C"
/codon_start=1
/product="Cdc25p: regulatory protein of adenylate cyclase"

```

```

/ protein_id="AAB64528.1"
/db_xref="pid:g2258166"
/translation="MSDTNTSIPNTSSAREAGNAQPSISSSSNTSTTTNTSSAS
LSSPSTSELSTIRPIGIVAAVDFNPPIKKSSQLSVQGGTYILKNSGMD
GLYIDNSGNVNGWPFONFGRPLROSHLRKSHGPMKYSKSSRRSLKNSAY
LHPRPSKRGSSSTLSASLNHNAETSSGHNNTYMANNSPSSAPNDASHITPOS
NENSNASLSODMTKRSADSSSEMTNAINNNETNLOJSGERAGPVAEBTITLPLE
ETEMINGIRSNASTWSPILITKTSIDYKLVYKNDLITYCSELPISNIMESDI
CNDEKPPPPNDHLVNLITRDLRNNANIEDSTSKQSESDQNRSLMERKQSDTGD
NNSINDDNHNNNNKNEFNEAGPSSLSAPDLQNSRYVAPSRSSILAKSDIF
YHSRDKLWTELDLTYVTYTKAHKMLKENLNFKYFDLISDVIPTOLGCRRL
HEIKASCSCKEIKKIFKGLISSISRSINSHLYFTDSAPHRKMDTAMDKNNDENN
SRTEGDGKLEIVSDHDLVSPJSGKRNVTSTDTITPKRSFTVENDEMFVSL
GPRNSVNSVTPRTSIOSTLEDFSPSNKPKSAKSYEMADVEF"
/ complement(2279. .2626)
/ gene="YLR311C"
/ complement(2279. .2626)
/ gene="YLR311C"
/ codon_start=1
/ evidence=not_experimental
/ product="Ylr311cp"
/ protein_id="AAB64518.1"
/db_xref="pid:g662134"
/db_xref="GI:662134"
/translation="MKLTKEKNDCLVGVYIPLNFTLFLFLRIEYVHSLIS
LSLSLRYFHNVCYPSLFLFCFVPIPFYSVRILLYLILRSFYELNILLYGAN
SRQSPGYVIR"
/ complement(2651. .3847)
/ gene="YLR312C"
/ complement(2651. .3847)
/ gene="YLR312C"
/ codon_start=1
/ evidence=not_experimental
/ product="Ylr312cp"
/ protein_id="AAB64517.1"
/db_xref="pid:g662133"
/db_xref="GI:662133"
/translation="MSEEDHNLVLRILRKRGEGEGSSKSEISLDSLHSSPAG
DDEDFDADVLNNTSSESAQMNRIYDRTSNEFSNAGVNDQGVPTISSEFDLSGS
NMGTVLPMEGSLKNDSTIRNSSTLSDHIDKSEGSAKIKMHVYIMLSLMTGS
YLAETVLTGDLVAGFKSOOSLRNNEKLLYGNIDPYDKSSYSSDSLSOMASGY
YVDEDNIAIAPSLRDLDMGRRYKTDVILMYTKAMKDMGRIRIKINGRKKL
FLNSTRSAOESLARVLHKEOKRRKKRLVTLNKKYRQPSPIKRTIFHSCOKAKOCS
GSRLOQLKRLFKSNKPPRFVFRKDTMFWQLKRPGLKLOHSRYKAMSECRKN
YFKCKH"
/ 4494. .5255
/ gene="MRPL15"
/ 4494. .5255
/ gene="MRPL15"
/ note="Ylr312w-ap"
/ codon_start=1
/ product="Mrpl15p: mitochondrial ribosomal protein yml15"
/ protein_id="AAB64533.1"
/db_xref="pid:g2258171"
/db_xref="GI:2258171"
/translation="MENSMAEISHSILRPYALNCNQSVRTVYILKGRINGLRD
PESTLRNPSCVLTLEVNAKECODVRSIILDPKYGIMLSNELLLOCLTHKSFAGSAP
YNEKLNLGAOFKLOTCIHSKNGSPAESCENGLQSFNSLGTFRPAKELTSKNTAC
TFVRLHLDPEIFMKMRDPKIDGHINGETTFASVLAFAIGALISTGSKAAKFIOG
SLDKEDLHSLVPIANENVASAKAKISDKENKAFV"
/ complement(5764. .7356)
/ gene="YLR313C"
/ complement(5764. .7356)
/ gene="YLR313C"
/ note="Note that a frameshift between nucleotides
5765-5805 would extend this gene. However, no mistake in
the sequence is apparent"
/ codon_start=1
/ evidence=not_experimental
/ product="Ylr313cp"
/ protein_id="AAB64516.1"
/db_xref="pid:g662132"

```

```

/db_xref="GI:662132"
/translation="MISSELTNDQIIDLSDYKRRKRIIASVPEDDRNRONRN
KDLTKLSNVQWLTLDVNDLMLKRLTDSADASLNDLDFEKKAKOESLRDLDAKH
KLILDIETBEIERNLHLDGNTNNGDEDLNDFYNDTLEFKEIIDDMPVNGIIS
IEVERELKTOFTLYFQNTLRIDPDTTTRLLILETTIKIKLIGDLDPVLSYL
QSILENTELYLAKASHAITNSTEPLVYGLDIPRIYAOSSISEVTFAPGAIQVTK
STRDOLSRNEGSLDLEAMKRLKTEYKRNKNGDISLDSGCTAPSPSSGCP
ITRKSDPVPVVASPSISITIKSSSISRESGVKRNSTSGETNLASVSLATKNSRIT
SEPSRGLGLKAVNSNSPSPQNTPLPLGKFRQDQASPPKRVITKPAETAKPAPA
NIPPAADVLSPTVTKMRKFRERYOKFAPNSGLRLISTEENLNNSDVSTHMANI
NNVFEVSSSNVYLPMAGRAF"
/ complement(7589. .9151)
/ gene="CDC3"
/db_xref="SGD:S0004306"
/ complement(7589. .9151)
/ gene="CDC3"
/ note="YLR314C: cell division cycle protein"
/ codon_start=1
/db_xref="SGD:S0004306"
/ product="Cdc3p: component of 10 nm filaments of
mother-bud neck"
/ protein_id="AAB64515.1"
/db_xref="pid:g662131"
/db_xref="GI:662131"
/translation="MSLKEQVSTIKOPDEEERHOQFNDOVTKOESODHDGYSOYT
NGTNDSEKFEAEASDVKVEPLGLGKITSSOSKQVLDQPEIKITRQINVTGF
ANLPKQWHRRSIRNGFENLVCVPPDIGITLTKLTFENNDDIEANLVKQIEELAD
QEEBEGQGEHENSQOQRKRVKISYESTIENGVLNINVIDTEGDFGLNNDQS
MDPIKEIDRSPOYLDAENKRNHSINDRIHACLYIEPTGHLPLDKFQWQSY
EKNCLIPVIAKSDILDEITLSEFKTIMNOLIOSNIEFLFPPIYSNDANSHSEEL
FSSLPVAVTNSNDIVENYSGNVGRGSPGVGLYENDNSDFLKNLILKOMFEL
KERTSKLTENYSSKLAKGIRKODNSVFREFDPIQOLEKTIHEKTLKLEIEMKT
VFOQVSEKERRKLOKSETLFAHREKREKLVQLKALEKKOLELISNASAPVNH
SPVTKKKGFLR"
/ 9822. .10283
/ gene="YLR315W"
/ 9822. .10283
/ gene="YLR315W"
/ codon_start=1
/ evidence=not_experimental
/ product="Ylr315wp"
/ protein_id="AAB64519.1"
/db_xref="pid:g662135"
/db_xref="GI:662135"
/translation="MNSEQDLHNVSDLLTTLISFQEFKQQLQSYTSDEQLOHWE
LQARDARVSELEARIKQFTILRSLLFLFEQSHLSLETLLDALYKINDLQ
ORLQILDDAIOERTSELAERENWVRSPSAGDNAIPGLQIIQSYINLEEN"
/ complement(10280. .10771)
/ gene="YLR316C"
/ complement(10280. .10771)
/ gene="YLR316C"
/ codon_start=1
/ product="Ylr316cp"
/ protein_id="AAB64529.1"
/db_xref="pid:g2258167"
/db_xref="GI:2258167"
/translation="MDEVNRELRSASTLSYKMATAGKQPMVSVFVDPSSRKDKVVAE
DGRNCENSLIDHSYVNGIRAVGRLEGVDEADNSYLCIDYVTLHEGCSKSMAL
IHSVVRVIVLEMQFRGSLKINSGDGYCMNDNQINSYTEAFQIGEEYVPGQVDRD
VCC"

```

Query Match 83.2%; Score 15.8; DB 8; Length 37730;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGATCAGACATGCT 19
||||| |||||||||
DB 10864 AAGGAATGAAACATGCT 10846

RESULT 13
HS426N21/c 44826 bp DNA PRI 08-DEC-1998
LOCUS

DEFINITION Human DNA sequence from clone 426N21 on chromosome X Contains EST, STS, GSS, complete sequence.

ACCESSION 282208

NID 93980421

VERSION 282208.1 GI:3980421

KEYWORDS HMG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 44826)

AUTHORS Grafham D.

JOURNAL Direct Submission

Submitted (07-DEC-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

On Dec 8, 1998 this sequence version replaced gi:1772964.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

IMPORTANT: This sequence is not the entire insert of clone 426N21. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true right end of clone d537K23 is at 100 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/RGP/CHRX>

426N21 is from the library RPI3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://dacpac.med.buffalo.edu/VECTOR: pcrPAC2>.

FEATURES

location/Qualifiers

1. 44826

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/clone="426N21"

/clone_1lb="RPI3"

39..350

repeat_region

/note="AluX repeat: matches 1..312 of consensus"

misc_feature

<1386..>1723

/note="match: GSS AQ131901"

2516..2557

/note="MER5A repeat: matches 146..189 of consensus"

2556..2641

/note="MIR repeat: matches 28..116 of consensus"

2642..2944

/note="AluY repeat: matches 4..297 of consensus"

2954..3268

/note="AluX repeat: matches 1..312 of consensus"

3269..3282

/note="MIR repeat: matches 124..137 of consensus"

3283..3587

/note="AluX repeat: matches 1..305 of consensus"

3588..3720

/note="MIR repeat: matches 137..262 of consensus"

3733..4032

/note="AluY repeat: matches 1..301 of consensus"

4425..4605

/note="AluSg/x repeat: matches 131..311 of consensus"

4606..4921

/note="AluSg repeat: matches 1..313 of consensus"

4922..4958

repeat_region

/note="AluSg/x repeat: matches 95..131 of consensus"

4961..5008

/note="AluSg/x repeat: matches 129..185 of consensus"

5009..5142

/note="AluO/FlAM repeat: matches 2..185 of consensus"

5151..5464

/note="AluSg repeat: matches 2..307 of consensus"

5467..5656

/note="AluY repeat: matches 119..305 of consensus"

5657..5969

/note="AluX repeat: matches 1..299 of consensus"

6219..6357

/note="AluX repeat: matches 1..141 of consensus"

6906..7204

/note="AluX repeat: matches 1..298 of consensus"

7288..7587

/note="AluSg repeat: matches 1..309 of consensus"

7595..7694

/note="MIR repeat: matches 107..223 of consensus"

8309..8807

/note="AluYB repeat: matches 1..305 of consensus"

8939..9241

/note="AluX repeat: matches 1..304 of consensus"

10115..10418

/note="AluX repeat: matches 1..306 of consensus"

11014..11319

/note="AluX repeat: matches 1..308 of consensus"

11483..11512

/note="15 copies 2 mer at 87% conserved"

12044..12121

/note="MER5A repeat: matches 85..162 of consensus"

12348..12577

/note="15 copies 2 mer tt 87% conserved"

12800..13093

/note="AluX repeat: matches 17..311 of consensus"

13197..13475

/note="AluX repeat: matches 1..281 of consensus"

13548..13839

/note="AluSg repeat: matches 1..300 of consensus"

13979..14371

/note="LTR37A repeat: matches 7..426 of consensus"

14397..14440

/note="22 copies 2 mer tg 89% conserved"

14498..14579

/note="12 repeat: matches 2635..2710 of consensus"

14580..14890

/note="AluSg repeat: matches 1..313 of consensus"

14891..16029

/note="12 repeat: matches 1456..2635 of consensus"

16038..16091

/note="18 copies 3 mer gag 74% conserved"

16570..16861

/note="12 repeat: matches 2288..2517 of consensus"

16862..17172

/note="AluX repeat: matches 1..313 of consensus"

17173..17369

/note="12 repeat: matches 2517..2731 of consensus"

17744..18046

/note="AluX repeat: matches 1..303 of consensus"

19119..19431

/note="AluX repeat: matches 1..299 of consensus"

19861..20160

/note="AluY repeat: matches 1..301 of consensus"

20205..20371

/note="AluO repeat: matches 1..169 of consensus"

21016..21331

/note="AluX repeat: matches 1..312 of consensus"

21318..21413

/note="Single clone region, single clone in alu"

21429..21721

/note="AluSg repeat: matches 1..295 of consensus"

21767..22064

/note="AluSg repeat: matches 1..299 of consensus"

repeat_region 22570..22639
/note="12 repeat: matches 2430..2458 of consensus"
repeat_region 22685..22984
/note="AluSP repeat: matches 1..306 of consensus"
repeat_region 24345..24535
/note="AluSG/x repeat: matches 128..312 of consensus"
repeat_region 24729..25043
/note="AluX repeat: matches 1..311 of consensus"
prim_transcript complement(26632..>26963)
/note="match: 3' EST A1087196 clone IMAGE:1679850"
repeat_region 26893..27480
/note="14 copies 42 mer 59% conserved"
repeat_region 27062..27493
/note="9 copies 48 mer 63% conserved"
repeat_region 27305..27516
/note="106 copies 2 mer cc 55% conserved"
repeat_region 30587..30660
/note="Charle4 repeat: matches 1882..1956 of consensus"
repeat_region 30935..31131
/note="MIR repeat: matches 1..206 of consensus"
repeat_region 31143..31437
/note="AluSG repeat: matches 1..299 of consensus"
repeat_region 31553..31744
/note="MIR repeat: matches 7..162 of consensus"
repeat_region 31745..32000
/note="LIMB7 repeat: matches 5923..6173 of consensus"
repeat_region 32001..32089
/note="MIR repeat: matches 162..259 of consensus"
repeat_region 32690..32767
/note="AluS repeat: matches 1..78 of consensus"
repeat_region 33718..33857
/note="MIR repeat: matches 82..216 of consensus"
repeat_region 35288..35587
/note="AluB repeat: matches 1..296 of consensus"
repeat_region 35738..36047
/note="AluX repeat: matches 2..311 of consensus"
repeat_region 38259..38441
/note="MIR repeat: matches 8..192 of consensus"
repeat_region 38748..38774
/note="9 copies 3 mer gag 93% conserved"
repeat_region 38748..38774
/note="9 copies 3 mer gag 93% conserved"
repeat_region 41141..41332
/note="match: GSS A0000811 clone 2282A5"
repeat_region 43144..43132
/note="MIR repeat: matches 60..259 of consensus"
repeat_region 43144..43454
/note="AluX repeat: matches 2..312 of consensus"
repeat_region 43851..44046
/note="MIR repeat: matches 25..219 of consensus"
misc_feature complement(44587..44786)
/note="match: STS A1008917 clone 426N21"
BASE COUNT 10821 a 10536 c 11376 g 12273 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 9; Length 44826;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGATCAGACATGCCT 19
||||| ||||| |||||
DB 40741 AAGATCAGACATGCCT 40723

RESULT 14
LOCUS HUMTPII/c 4446 bp mRNA PRI 07-MAR-1995
DEFINITION Homo sapiens tripeptidyl peptidase II mRNA, 3' end.
ACCESSION M55169 J05299
NID 9339877
VERSION M55169.1 GI:339877
KEYWORDS tripeptidyl peptidase II,
SOURCE Human B-lymphocyte cell line SPL, and fibroblast cell line AG 1523,
CDNA to mRNA, clones A2 and F5.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4446)
AUTHORS Tomkinson,B. and Jonsson,A.K.
TITLE Characterization of cDNA for human tripeptidyl peptidase II: the
N-terminal part of the enzyme is similar to subtilisin
JOURNAL Biochemistry 30 (1), 168-174 (1991)
MEDLINE 91105077
FEATURES
source location/Qualifiers
1..4446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="A2"
/cell_type="SPL"
/cell_type="B lymphocyte"
/tissue_id="lambda gt10"
/clone="F5"
/cell_line="AG 1523"
/cell_type="fibroblast"
/tissue_id="lambda gt11"
1..3585
/gene="tripeptidyl peptidase II"
1..3585
/partial
/gene="tripeptidyl peptidase II"
/codon_start=1
/product="tripeptidyl peptidase II"
/protein_id="AA63263.1"
/db_xref="PID:9339878"
/db_xref="GI:339878"
/translation="VTIDGPKRIYDIIDTGGSDVNTAEVEPKDGLVIGSGRYLKI
PASWTNPSGKHIGIKNGYDYPKALKRIKEREKIMDVHVRALAEARCOEED
VANNGSOANKLKEELOSOVELNLSPEKYSDDPVDCLVHMGDEYVRAICIDNE
GDLSEKTVLANKKAEQSGSFGTAELMNSVNIYDNGULSLVSGAHGTHVSI
GHPEPERNGVAREQALISIKIGTRLSMETGGLIRAHIEVNHKCDLVNYSYGE
ATHNPSGRICEVINEAYKKNHITVVSAGNNGPCLSTVGCGGTTSSVIGAYVSP
DMVAEYSLREKLPANQYTWSSRGSAGALGVSLISAGGAIASVPNTLTGQVLMNG
TSMSPNAGGIALITLSGLKANNIDYVHVSRALENAVADNIEVFAOCHGILQVD
KAYDVLQNTSEFANKLGGTYVGNRGITLDPVAVAPSDHGVGIEVEPENTSE
KISLOHLALTNSSWOCPSHLELMNCRHINIDVDRGLREGIHYEVGGYDIASP
NAGPLFRYPITAVIAAKNBSHSHLATVDVHFRGQIRRHIEVEPEATVAEYVCS
CSSEYSAFVLAHVLVOKAIRSHHEFCSLPKGLTEHFPVLSGKALPCLAKM
MASLSDVNIIDYTIHFGITCAPOLNHAISGINSINEFDOSSILKEDLAPCTLLKMWQ
TLRPVSATKPLGSRDVLPPNNKOLYEMALITFHPKSGEYTPSCPLCELYESEFD
SOLMTIIPDNKNGSGDAYPHOYSLKLEKDYITRLDIRHOISDLRLKDLPIVS
HRLSNTLSLDIHENHSFALDKKSSNLTLPKYNQPEFVSLPDKIPKGGPCYL
AGSLTSTKEIGKRDVLPVHYLLPPTTKRNGSKDEKSEKDKLEPTEALRD
LKIDMTKIDSDIYNELKETYPNTLPVVARLHOLDKEKEMKRLNETVDANAVIS
HIDQTLAVYIEMAKTDPDDAATIKNDKQKSTLVDAICRGKALADHLHTQADG
AISTDAEKKEEGESPLDSLAETFEWETKMDLLENKRVLTAYKHALVNMKYGRIKF
ATKVEEPTKEMNKNCIQMLKILMTWCASFTEWMLPIMPPDYCV"

BASE COUNT 1432 a 820 c 924 g 1270 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 9; Length 4446;
Best Local Similarity 89.5%; Pred. No. 85;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGATCAGACATGCCT 19
||||| ||||| |||||
DB 1224 AAGATCAGACATGCCT 1206

RESULT 15
LOCUS HUMTPII/c 4626 bp mRNA PRI 07-AUG-1992
DEFINITION Homo sapiens tripeptidyl peptidase II mRNA, complete cds.
ACCESSION M73047 J05299 M55445 M72378
NID 9339879
VERSION M73047.1 GI:339879
KEYWORDS tripeptidyl peptidase II.

SOURCE Human, cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 4626; 889 to 4626)
Tomkinson, B. and Jonsson, A.-K.
Characterization of cDNA for human tripeptidyl peptidase II: The
N-terminal part of the enzyme is similar to subtilisin
JOURNAL Biochemistry 30, 168-174 (1991)
MEDLINE 91105077
REFERENCE 2 (bases 1 to 4626; 1 to 888)
Tomkinson, B.
Nucleotide sequence of cDNA covering the N-terminus of human
tripeptidyl peptidase II
JOURNAL Biomed. Biochim. Acta 50, 727-729 (1991)
MEDLINE 92198394
FEATURES
location/Qualifiers
1..4626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="SPL"
/cell_type="B-lymphocyte"
/tissue="lb="lambda gt10"
24..3773
/codon_start=1
/product="tripeptidyl peptidase II"
/protein_id="AA36760.1"
/db_xref="GI:339880"
/db_xref="PI:9339880"
/translation="MATATEPEEPFPHGLPKKETGAASFLCRYPEYDGRGLIAVLD
TGVDGAPGMOVTIDGPKIVDIDTSGDVNTATPEVDEIGVLSGRVLKIPAS
WNPSCGKHYHIGKNGYDFPKAKERIQKREKIMDPYHVALAECKROEEDVAN
NGSSQAKLIREELQSVLELNSFEKISDPGVYDCLVHDEWVACIDSNDL
SKSTVLANRYEAGQYSGFTGAEMLNSVNIYDDNLISITVSGAGTHVSAIAAGH
PEEPERGVAPGAQILSIKIGDRLSTMETGILIRAMIEVINHKCDLVNYSYEATH
WPNSGRICEVINAEVWKHNIIVYSAGNNGCLSTVCGPGGTSVVGAVSPDM
VAEYSLREKLPANQYTWSSSRGSPSADGALGVISAPGALASVPNMTLRGOLMGTS
SSPNACGGAILLSGLKANNIDTVHSHVRALENTAYKADNIEVPAOGRGCIQYDKAY
DYLVQNTSFANKLGFVTVGNNGITLRDPOVAAPSDHGVGLEVPENTENSEKIS
LQHLALTSNSWVOCPSHLELNOCRHINIRVDPRGLRGLHYTEVCGDIASPNAG
PLFRVPTAVIAAKVNESHSHYDLFTVDFHFKPQIRRHFEVEGATMAEVYSCSS
EVSARFVIAVAVOKORAYRSHFEYFKCS.PEKGTLAEAPVVGKALIEFCIARWAS
LSDVNIDYTSFPHIVCTAPOLNHAISEGINRPDVOSLSKVEDLAPCITLKNVOTLR
PVSAKTRPLGSRVPLPNNKRLYEMVLTYNHPKSGEYTPSCPLLCELYESEFDSOL
WIFDOKRQWGSQDAIPHQYSLKLEKGDYTRIQIRHEQISDLERKDLPTVSHRL
SNTLSLDIHENHSFALLGKRKSNLTLPYNOPTFYVSLPDKIPRGAPGCYLAGS
LTLSTELGRKAVIPVHYLLIPPTKNGSKDEKDESEKDLKEEFTALDLKI
QWMTKLDSDIYNELKETYPNYLPLVYARLHOLDAEKERRKRLNEIYDANAVYSHD
QVALAVYIAKMTDPRPDATIKNDMDKOKSTLVDAICRGCALADHILHTNOADGALS
TDAEGKEESESPLDSLAETFTETTKTDTLFDNKVULTEFAKHALVNMKMGRLKFAIK
LVEKEPTKEMKNCIQIMKLLGWTHCASTENWLPIMYPDYCVF"

BASE COUNT 1452 a 885 c 983 g 1306 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 9; Length 4626;
Best Local Similarity 89.5%; Pred. No. 85;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGATCAGACATGCCT 19
|||||||
Db 1412 AAGATCAGGCAATGCCT 1394

Search completed: September 13, 1999, 15:55:53
Job time: 4581 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

CH nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:34 ; Search time 255.05 Seconds

(without alignments)
18.638 Million cell updates/sec

Title: US-09-325-095-24

Perfect score: 19

Sequence: 1 AAGGATCGAGACAAATGCCCT 19

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15.8	83.2	7146	1	V38933	Nucleic acid seque
2	15.8	83.2	17425	1	X28366	Human Stat6 gene.
3	14.8	77.9	1987	1	N80778	CDNA sequence for
4	14.8	77.9	17410	1	O53142	Sequence encoding
5	14.8	77.9	218	1	O76820	Human genome fragm
6	14.8	77.9	17350	1	O72708	hnp1 human osteoge
7	14.8	77.9	17415	1	T11639	Human osteogenic p
8	14.8	77.9	17410	1	T18381	hop-1 genomic DNA.
9	14.8	77.9	3931	1	T58758	Murine long form C
10	14.8	77.9	1987	1	T58758	Murine long form C
11	14.8	77.9	8543	1	T47073	Maize dwarf mosaic
12	14.8	77.9	1296	1	T47074	Maize dwarf mosaic
13	14.8	77.9	3931	1	T80599	Murine 4 kb colony
14	14.8	77.9	1987	1	T80600	Murine 2 kb colony
15	14.8	77.9	3931	1	T94752	Murine colony stim
16	14.8	77.9	1987	1	T94753	Murine colony stim
17	14.8	77.9	3931	1	V01083	Gene for mouse col
18	14.8	77.9	1987	1	V01084	Gene for mouse col
19	14.8	77.9	17410	1	V15205	Human osteogenic p
20	14.8	77.9	137507	1	V19941	KSHV long unique c
21	14.8	77.9	979	1	V81624	Arbidiopsis PR-R p
22	14.8	77.9	35100	1	V73804	KSHV LTR DNA (nucl
23	14.8	77.9	553	1	V84417	Human secreted pro
24	14.8	77.9	979	1	V82844	Arbidiopsis PR-5 g
25	14.8	77.9	17415	1	V80733	Human osteogenic p
26	14.8	77.9	17410	1	X00230	Murine PGC-1 CDNA.
27	14.8	77.9	3066	1	X02188	Polynucleotide seq
28	14.8	77.9	3219	1	X20565	Human brain expres
29	14.4	75.8	433	1	O59567	Streptococcus pneu
30	14.4	75.8	10974	1	V53247	EST clone GW707. N
31	14.4	75.8	624	1	V85990	EST clone BG457. N
32	14.4	75.8	311	1	V86701	Attenuated hepatit
33	14.2	74.7	7493	1	N00512	Sequence encoding
34	14.2	74.7	2837	1	N00512	Murine serotoiner
35	14.2	74.7	2036	1	O70264	Proteasom CDNA. S
36	14.2	74.7	2740	1	O85355	Delux CDNA. Novel
37	14.2	74.7	3771	1	O92980	Delux locus compo
38	14.2	74.7	5063	1	O92979	CDNA encoding huma
39	14.2	74.7	4165	1	T16483	Drosophila deltex
40	14.2	74.7	3771	1	T68924	Hepatitis A virus
41	14.2	74.7	7493	1	T93023	Human eps15 protei
42	14.2	74.7	4165	1	V13998	Hepatitis A virus
43	14.2	74.7	7494	1	V05898	Hepatitis A virus

ALIGNMENTS

RESULT 1	
ID V38933/C	standard: DNA: 7146 BP.
AC V38933:	
DT 24-SEP-1998	(first entry)
DE Nucleic acid sequence of genomic DNA encoding human STCP-1.	
KW Human; STCP-1; chemokine activity; T-cells; treatment; HIV infection;	
KW Inhibitory compound; assay; reduce; circulatory system STCP-1 level;	
KW Joint inflammation; rheumatoid arthritis; lupus; ds.	
OS Homo sapiens.	
FT Key	Location/Qualifiers
FT CDS	1437..6204
FT	/*tag- a
FT	/product= STCP-1
FT	/note= "contains introns and exons"
FT	1437..1509
FT	/*tag- b
FT	/number= 1
FT	1510..3056
FT	/*tag- c
FT	/number= 1
FT	3057..3180
FT	/*tag- d
FT	/number= 2
FT	3181..6119
FT	/*tag- e
FT	/number= 2
FT	6120..6201
FT	/*tag- f
FT	/number= 3
PM W09624907-A1.	
PD 11-JUN-1998.	
PR 26-NOV-1997; U21552.	
PR 03-DEC-1996; US-760127.	
PA (AMGE-) AMGEN INC.	
PI Andrew DP; Chang M;	
DR WPI: 98-333326/29.	
DR P-PSDB; W62783.	
PT Human STCP-1 polypeptides with chemokine activity - useful e.g. to	
PT treat HIV infection or other viral or bacterial pathogens infecting	
PT T-cells, macrophages or other immune system cells	
PS Claim 1; Fig 2A-F; 96pp; English.	
CC The present sequence encodes human STCP-1. STCP-1 polypeptides	
CC demonstrate chemokine activity for T-cells. The polypeptides are useful	
CC prophylactically or therapeutically to treat HIV infection and other	
CC conditions associated with viral/Bacterial pathogens infecting T-cells,	
CC macrophages or other immune system cells. They can be included	
CC (optionally chemically modified) with a pharmaceutically acceptable	
CC carrier and optionally other pharmaceuticals (e.g. AZT, antibiotics etc.)	
CC in therapeutic compositions for treating these conditions. STCP-1 also	
CC useful to assay for inhibitory compounds used to reduce circulatory	
CC system STCP-1 levels to alleviate e.g. joint inflammation associated	
CC with rheumatoid arthritis, lupus or other autoimmune diseases. The	
CC polypeptides are also useful to prepare antibodies or hybridomas. The	
CC nucleic acids are useful to produce hybridisation probes to test for	
CC STCP-1 DNA/RNA in mammalian samples.	
SQ Sequence 7146 BP; 1664 A; 1814 C; 1883 G; 1785 T;	

Query Match 83.2%; Score 15.8; DB 1; Length 7146;
Best Local Similarity 89.5%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAGGATCGAGACAAATGCCCT 19
DB --4036 AATGATCGAGACAAATGCCCT 4018

```

RESULT 2
X28366/C
ID X28366 standard; DNA; 17425 BP.
AC X28366;
DE 18-JUN-1999 (first entry)
KW Human Stat6 gene.
KW Stat6; Stat6b; human; signal transducers and activators of transcription;
KW isoform; myeloid cancer; asthma; sarcoma; scleroderma; fibrotic disease;
KW bone marrow fibrosis; AIDS; Stat6c; ss.
OS Homo sapiens.
PN M09310493-A1.
PD 04-MAR-1999.
PE 27-AUG-1998; 017821.
PR 05-JAN-1998; 05-070397.
RA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Larochelle WJ, Patel B, Pierce JH;
DR WPI; 99-214517/18.
PT New isoforms of Stat6 - having differential effects on the
PS modulation of Stat6 activity in cells.
CC This sequence encodes human Stat6 (signal transducers and activators
CC of transcription). The invention relates to attenuated and dominant
CC negative isoforms of human Stat6. The detection and quantitation of DNA
CC or mRNA encoding Stat6 and/or Stat6b and/or Stat6c can be used to detect
CC differential expression of Stat6 isoforms in numerous diseases, including
CC myeloid cancer, asthma, sarcoma, scleroderma, bone marrow fibrosis,
CC fibrotic diseases and AIDS. The nucleic acids can be used to screen
CC genomic or cDNA libraries or to identify complementary sequences. The
CC identification of the genetic locus of the Stat6 gene can be used for
CC detection of chromosomal aberrations and the Stat6 gene can be used for
CC Stat6 gene. Antibodies against the isoforms can be used to detect the
CC presence of Stat6 and/or Stat6b and/or Stat6c in a sample. Because of the
CC variation of the roles of Stat6b and Stat6c in regulating gene
CC transcription the isolated and purified forms can be used to study gene
CC regulation and in screening assays for identifying drug candidates which
CC may be used as agonists or antagonists. The two polypeptides may also be
CC used in gene therapy protocols. In particular, Stat6b and/or Stat6c can
CC therapeutically modulate the development and differentiation of B and
CC T cells and can enhance IL-4 immunological function in immunocompromised
CC individuals. Stat6 activation correlates with functional responses
CC induced by interleukin-4 (IL-4), IL-13 and platelet-derived growth factor
CC (PDGF). Stat6b when compared Stat6 is an attenuated regulator of gene
CC transcription. Stat6c is a dominant negative regulator of gene
CC transcription.
SQ Sequence 17425 BP; 3991 A; 4519 C; 4691 G; 4220 T;

Query Match 83.2%; Score 15.8; DB 1; Length 17425;
Best Local Similarity 89.5%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGATGAGAACATGCT 19
DB 1806 AAGGCTTGAACATGCT 1788
|||||

```

```

RESULT 3
N80778
ID N80778 standard; cDNA; 1987 BP.
AC N80778;
DE 15-OCT-1990 (first entry)
DE cDNA sequence for a murine cDNA 2kb clone encoding murine colony
DE stimulating factor-1 (mucSF-1)
KW Murine colony stimulating factor-1; 2kb clone; murine L-929.
OS Mouse
PN Mouse
FH Key
FT cds
FT Location/Qualifiers
FT 70..165
FT /*tag= a
FT /product=leader peptide
FT mat_peptide 166..1728

RESULT 4
Q53142/C
ID Q53142 standard; DNA; 17410 BP.
AC Q53142;
DE 09-JAN-1996 (revised)
DE 06-JUN-1994 (first entry)
DE Sequence encoding osteogenic protein.
DE Osteogenic protein; bone; cartilage; matrix; osteoarthritis;
KW repair; vascularisation; mineralisation; differentiation; ss.
OS Homo sapiens.
PN Key
FH Key
FT exon
FT Location/Qualifiers
FT 3192..3730
FT /*tag= a
FT /label= Exon 1.
FT /note= "Start codon begins at position 3313."
FT 10413..10414
FT /*tag= b
FT /label= Gap 1
FT /note= "Bases are estimated to be missing between
FT these two positions in this sequence."
FT 10696..10891
FT /*tag= c
FT /label= Exon 2.
FT 10960..10961
FT /*tag= d
FT /label= Gap 2
FT /note= "Bases are estimated to be missing between
FT these two positions in this sequence."
FT 11059..11211
FT /*tag= e
FT /label= Exon 3.
FT 11351..11352
FT /*tag= f
FT /label= Gap 3.
FT /note= "Bases are estimated to be missing between
FT these two positions in this sequence."
FT 11420..11517
FT /*tag= g
FT /label= Exon 4.

Query Match 77.9%; Score 14.8; DB 1; Length 1987;
Best Local Similarity 88.9%; Pred. No. 73;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGATGAGAACATGCT 19
DB 452 AAGAGCGAAGACATGCT 469
|||||

```

```

FT misc_feature 11721. 11722
FT //tag= b
FT /label= Gap 4.
FT /note= "Bases are estimated to be missing between
FT these two positions in this sequence."
FT FT exon 13354. 13436
FT //tag= i
FT /label= Exon 5.
FT FT exon 15044. 15160
FT //tag= j
FT /label= Exon 6.
FT FT exon 17245. 17410
FT //tag= k
FT /label= Exon 7.
PN US5266683-A.
PD 30-NOV-1993.
PE 08-APR-1988; 179406.
PR 08-APR-1988; US-179406.
PR 15-AUG-1988; US-232630.
PR 23-FEB-1989; US-315342.
PR 17-OCT-1989; US-422613.
PR 17-OCT-1989; US-422699.
PR 22-FEB-1990; US-483913.
PR 20-AUG-1990; US-569920.
PR 07-SEP-1990; US-579865.
PR 18-OCT-1990; US-599543.
PR 18-OCT-1990; US-600024.
PA 04-DEC-1990; US-621849.
PA 04-DEC-1990; US-621988.
PA 22-FEB-1991; US-660162.
PA 20-DEC-1991; US-810560.
PA 28-JAN-1992; US-827052.
PA 21-FEB-1992; US-841646.
PI Kuberassampath T, Oppermann H, Ozkaynak E, Pang RHJ;
PI Rueger DC;
PI WPI: 93-395405/49.
DR P-PSDB: R44746.
PT New pure mammalian osteogenic proteins - induce cartilage and
PT endochondral bone formation when in association with a matrix
PT Disclosure; Columns 73-88; 128pp; English.
CC The osteogenic protein when in association with a matrix can
CC at the locus of an implant the full development cascade of
CC endochondral bone formation including vascularisation.
CC mineralisation and bone marrow differentiation. The osteogenic
CC protein can also be used to repair both bone and cartilage in the
CC treatment of osteoarthritis.
SO Sequence 17410 BP; 3833 A; 4574 C; 4718 G; 4285 T;

Query Match 77.9%; Score 14.8; DB I; Length 17410;
Best Local Similarity 88.9%; Pred. No. 94;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGCATCAGACAAATGCC 18
DB 1312 AAGGACGAGAAAATGCC 1295

RESULT 5
V/overl
ID Q76820 standard; DNA: 218 BP.
AC Q76820:
DT 23-SEP-1994 (first entry)
DE Human genome fragment.
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
OS Homo sapiens.
PN WO9401548-A.
PD 20-JAN-1994.
PE 13-JUL-1993; G01467.
PR 13-JUL-1992; GB-014657.
RA (MEDT-) MEDICAL RES COUNCIL.
PL Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;

```

```

PI      Stpson DR:Starkey M;
DR      WPI: 94-035056/04.
PT      New nucleic acid fragment encoding gene products - can be used
PS      for genetic analysis and mapping.
PS      Claim 1; Page 22; 616pp; English.
CC      Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC      the placenta or bone marrow comprise any of: (A) a sequence
CC      selected from (Q76401-Q77613), (B) an allelic variation of a
CC      sequence as described in (A), or (C) a sequence complementary
CC      to (A) or (B).
SQ      Sequence 218 BP; 45 A; 68 C; 55 G; 50 T;

Query Match      77.9%; Score 14.8; DB 1; Length 218;
Best Local Similarity 88.9%; Pred. No. 56;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAGCATCAGAACATGCC 18
        ||||| ||| |||||
DB      94 AAGCATCAGATCAATGCC 111

RESULT      6
ID      Q72708 standard; DNA: 17350 BP.
AC      Q72708:
DE      09-JUN-1995 (first entry)
KW      hox1: human osteogenic protein 1 partial genomic DNA.
KW      hox1: human osteogenic protein; osteoarthritis; orthogenesis;
KW      non-union fracture repair; allograft repair;
KW      cartilage and endochondrial bone formation;
KW      periodontal, dental and craniofacial reconstruction; ss.
OS      Homo sapiens
FH      Key
FH      exon
FT      Location/Qualifiers
        /tag_a
        /label=EXON-1
        10413..10414
        /tag_b
        /label=EXON-1
        /note="a gap occurs between positions 10413 and 10414
        in this sequence"
        10696..10891
        /tag_c
        /label=EXON-2
        10960..10961
        /tag_d
        /label=EXON-2
        /note="a gap occurs between positions 10960 and 10961
        in this sequence"
        11059..11211
        /tag_e
        /label=EXON-3
        11351..11352
        /tag_f
        /label=EXON-3
        /note="a gap occurs between positions 11351 and 11352
        in this sequence"
        11420..11617
        /tag_g
        /label=EXON-4
        11721..11722
        /tag_h
        /label=EXON-4
        /note="a gap occurs between positions 11721 and 11722
        in this sequence"
        13354..13436
        /tag_i
        /label=EXON-5
        15044..15160
        /tag_j
        /label=EXON-6
        17245..17410
        /tag_k
        /label=EXON-6
        17410..17411
        /tag_l
        /label=EXON-6
        17411..17412
        /tag_m
        /label=EXON-6
        17412..17413
        /tag_n
        /label=EXON-6
        17413..17414
        /tag_o
        /label=EXON-6
        17414..17415
        /tag_p
        /label=EXON-6
        17415..17416
        /tag_q
        /label=EXON-6
        17416..17417
        /tag_r
        /label=EXON-6
        17417..17418
        /tag_s
        /label=EXON-6
        17418..17419
        /tag_t
        /label=EXON-6
        17419..17420
        /tag_u
        /label=EXON-6
        17420..17421
        /tag_v
        /label=EXON-6
        17421..17422
        /tag_w
        /label=EXON-6
        17422..17423
        /tag_x
        /label=EXON-6
        17423..17424
        /tag_y
        /label=EXON-6
        17424..17425
        /tag_z
        /label=EXON-6
        17425..17426
        /tag_1
        /label=EXON-6
        17426..17427
        /tag_2
        /label=EXON-6
        17427..17428
        /tag_3
        /label=EXON-6
        17428..17429
        /tag_4
        /label=EXON-6
        17429..17430
        /tag_5
        /label=EXON-6
        17430..17431
        /tag_6
        /label=EXON-6
        17431..17432
        /tag_7
        /label=EXON-6
        17432..17433
        /tag_8
        /label=EXON-6
        17433..17434
        /tag_9
        /label=EXON-6
        17434..17435
        /tag_10
        /label=EXON-6
        17435..17436
        /tag_11
        /label=EXON-6
        17436..17437
        /tag_12
        /label=EXON-6
        17437..17438
        /tag_13
        /label=EXON-6
        17438..17439
        /tag_14
        /label=EXON-6
        17439..17440
        /tag_15
        /label=EXON-6
        17440..17441
        /tag_16
        /label=EXON-6
        17441..17442
        /tag_17
        /label=EXON-6
        17442..17443
        /tag_18
        /label=EXON-6
        17443..17444
        /tag_19
        /label=EXON-6
        17444..17445
        /tag_20
        /label=EXON-6
        17445..17446
        /tag_21
        /label=EXON-6
        17446..17447
        /tag_22
        /label=EXON-6
        17447..17448
        /tag_23
        /label=EXON-6
        17448..17449
        /tag_24
        /label=EXON-6
        17449..17450
        /tag_25
        /label=EXON-6
        17450..17451
        /tag_26
        /label=EXON-6
        17451..17452
        /tag_27
        /label=EXON-6
        17452..17453
        /tag_28
        /label=EXON-6
        17453..17454
        /tag_29
        /label=EXON-6
        17454..17455
        /tag_30
        /label=EXON-6
        17455..17456
        /tag_31
        /label=EXON-6
        17456..17457
        /tag_32
        /label=EXON-6
        17457..17458
        /tag_33
        /label=EXON-6
        17458..17459
        /tag_34
        /label=EXON-6
        17459..17460
        /tag_35
        /label=EXON-6
        17460..17461
        /tag_36
        /label=EXON-6
        17461..17462
        /tag_37
        /label=EXON-6
        17462..17463
        /tag_38
        /label=EXON-6
        17463..17464
        /tag_39
        /label=EXON-6
        17464..17465
        /tag_40
        /label=EXON-6
        17465..17466
        /tag_41
        /label=EXON-6
        17466..17467
        /tag_42
        /label=EXON-6
        17467..17468
        /tag_43
        /label=EXON-6
        17468..17469
        /tag_44
        /label=EXON-6
        17469..17470
        /tag_45
        /label=EXON-6
        17470..17471
        /tag_46
        /label=EXON-6
        17471..17472
        /tag_47
        /label=EXON-6
        17472..17473
        /tag_48
        /label=EXON-6
        17473..17474
        /tag_49
        /label=EXON-6
        17474..17475
        /tag_50
        /label=EXON-6
        17475..17476
        /tag_51
        /label=EXON-6
        17476..17477
        /tag_52
        /label=EXON-6
        17477..17478
        /tag_53
        /label=EXON-6
        17478..17479
        /tag_54
        /label=EXON-6
        17479..17480
        /tag_55
        /label=EXON-6
        17480..17481
        /tag_56
        /label=EXON-6
        17481..17482
        /tag_57
        /label=EXON-6
        17482..17483
        /tag_58
        /label=EXON-6
        17483..17484
        /tag_59
        /label=EXON-6
        17484..17485
        /tag_60
        /label=EXON-6
        17485..17486
        /tag_61
        /label=EXON-6
        17486..17487
        /tag_62
        /label=EXON-6
        17487..17488
        /tag_63
        /label=EXON-6
        17488..17489
        /tag_64
        /label=EXON-6
        17489..17490
        /tag_65
        /label=EXON-6
        17490..17491
        /tag_66
        /label=EXON-6
        17491..17492
        /tag_67
        /label=EXON-6
        17492..17493
        /tag_68
        /label=EXON-6
        17493..17494
        /tag_69
        /label=EXON-6
        17494..17495
        /tag_70
        /label=EXON-6
        17495..17496
        /tag_71
        /label=EXON-6
        17496..17497
        /tag_72
        /label=EXON-6
        17497..17498
        /tag_73
        /label=EXON-6
        17498..17499
        /tag_74
        /label=EXON-6
        17499..17500
        /tag_75
        /label=EXON-6
        17500..17501
        /tag_76
        /label=EXON-6
        17501..17502
        /tag_77
        /label=EXON-6
        17502..17503
        /tag_78
        /label=EXON-6
        17503..17504
        /tag_79
        /label=EXON-6
        17504..17505
        /tag_80
        /label=EXON-6
        17505..17506
        /tag_81
        /label=EXON-6
        17506..17507
        /tag_82
        /label=EXON-6
        17507..17508
        /tag_83
        /label=EXON-6
        17508..17509
        /tag_84
        /label=EXON-6
        17509..17510
        /tag_85
        /label=EXON-6
        17510..17511
        /tag_86
        /label=EXON-6
        17511..17512
        /tag_87
        /label=EXON-6
        17512..17513
        /tag_88
        /label=EXON-6
        17513..17514
        /tag_89
        /label=EXON-6
        17514..17515
        /tag_90
        /label=EXON-6
        17515..17516
        /tag_91
        /label=EXON-6
        17516..17517
        /tag_92
        /label=EXON-6
        17517..17518
        /tag_93
        /label=EXON-6
        17518..17519
        /
```

```

FT      US5354557-A.                                     /*tag- k
PD      11-OCT-1994.                                     /label= EXON-7
PF      08-APR-1988; US-179406.
PR      08-APR-1988; US-179406.
PR      15-AUG-1988; US-232630.
PR      23-FEB-1989; US-315342.
PR      17-OCT-1989; US-422613.
PR      17-OCT-1989; US-422699.
PR      22-FEB-1990; US-483913.
PR      20-AUG-1990; US-569920.
PR      07-SEP-1990; US-579865.
PR      18-OCT-1990; US-599543.
PR      18-OCT-1990; US-600024.
PA      04-DEC-1990; US-621849.
PA      04-DEC-1990; US-621888.
PA      22-FEB-1991; US-660162.
PA      20-DEC-1991; US-810560.
PA      28-JAN-1992; US-827052.
PA      21-FEB-1992; US-841646.
PA      18-DEC-1992; US-993387.
PI      Kuberansampath T, Oppertmann H, Ozkaynak E, Pang RHL;
PI      Rueger DC.
PI      WPI: 94-324521/40.
PT      Implantable device for inducing osteogenesis - comprises porous
PT      matrix contg. non-glycosylated dimeric, disulphide linked
PT      osteogenic protein.
PS      Disclosure: Columns 71-86: 128pp: English.
CC      072708 is the human osteogenic protein 1 (hOP1) genomic DNA sequence,
CC      the protein-coding region for hOP1 (R51644) is encoded in the seven
CC      exons. Fragments of this protein consisting of residues 335-431,
CC      318-431, 293-431, 300-431, 313-431, 315-431 and 316-431; are
CC      unglycosylated osteogenic polypeptides. Any two of these polypeptides
CC      can be disulphide bonded to form a dimer, which forms an essential
CC      component of an osteogenic protein. This protein is dispersed
CC      in a biodegradable matrix which can be implanted into a mammalian
CC      bone marrow cavity, here it can induce local cartilage, bone and
CC      endochondral bone formation; and it can also accelerate allograft
CC      repair. This implant has the advantage of inducing all stages of
CC      bone formation and of having a higher specific activity than other
CC      known biosynthetic materials. The implant can be used to repair
CC      non-union fractures and cartilage. Treat osteoarthritis; and aid
CC      in periodontal, dental or craniofacial reconstruction.
CC      Sequence 17350 BP; 3813 A; 4569 C; 4688 G; 4280 T;
SQ

```

Query Match	Best Local Similarity	Score	DB 1:	Length
Matches 16; Conservative	88.9%	77.9%	17415	0
1 AAGGATCAGAACATGCC 18	1314 AAGGAGCAGAAAAATGCC 1297			
expression changes in a cell transformed to express osteogenic protein-1 and having additional steroid binding site				
Claim 3: Page 43-53; 77pp; English.				
The human osteogenic protein-1 (OP-1) non-coding sequence is used in the construction of an expression vector comprising a reporter gene which has the non-coding sequence lying contiguous to the reporter gene and which is able to act on and affect the expression of the reporter gene when bound to by candidate compounds. The method is used to identify compounds capable of modulating OP-1 expression.				
Sequence 17415 BP: 3826 A; 4581 C; 4723 G; 4285 T;				
RESULT 8				
T18381/C				
ID T18381 standard; cDNA; 17410 BP.				
AC T18381;				
DT 20-JUN-1996 (first entry)				
DE hOP-1 genomic DNA.				
KW Human; osteogenic factor; hOP-1; murine; mOP-1; TGF-beta superfamily; transforming growth factor-beta; dimer; antibody; epitope; hippocampus; purification; implantable osteogenic device; bone formation; craniofacial; anomaly; skeletal; dental; endochondral bone formation;				
KW non-union fracture; cartilage repair; osteoarthritis; ss.				
OS Homo sapiens.				
PH Key				
FT exon	Location/Qualifiers			
FT	3192..3730			
FT	/*tag= a			
FT	/number= 1			
FT	3313..3315			
FT	/*tag= b			
FT	/function= Start codon			
FT	10413..10414			
FT	/*tag= c			
FT	/note= "Position of possible nucleotide deletion"			
FT	10697..10889			
FT	/*tag= d			
FT	/number= 2			
FT	11020..11021			
FT	/*tag= e			
FT	/note= "Position of possible nucleotide deletion"			
FT	11063..11210			
FT	/*tag= f			
FT	/number= 3			
FT	11351..11352			
FT	/*tag= g			
FT	/note= "Position of possible nucleotide deletion"			
FT	11421..11618			
FT	/*tag= h			
FT	/number= 4			
FT	11598..11599			
FT	/*tag= i			
FT	/note= "Position of deletion of G compared to cDNA sequence given in T02597"			
FT	11618			
FT	/*tag= j			
FT	/note= "Position of insertion of G compared to cDNA sequence given in T02597"			
FT	11721..11722			
FT	/*tag= k			
FT	13358..13435			
FT	/*tag= l			
FT	/number= 5			
FT	15047..15157			
FT	/*tag= m			

FT /number- 5
FT 17246. 17410
FT /tag- n
FT /number- 7
FT US5468845-A.
PN 21-NOV-1995.
PD 08-APR-1988; 179406.
PF 08-APR-1988; US-179406.
PR 15-AUG-1988; US-332630.
PR 23-FEB-1989; US-315342.
PR 17-OCT-1989; US-422699.
PR 22-FEB-1989; US-422613.
PR 20-AUG-1990; US-569920.
PR 07-SEP-1990; US-579865.
PR 18-OCT-1990; US-599543.
PR 21-NOV-1990; US-616374.
PR 04-DEC-1990; US-621849.
PR 04-DEC-1990; US-621988.
PR 22-FEB-1991; US-660162.
PR 20-DEC-1991; US-810560.
PR 28-JAN-1992; US-827052.
PR 21-FEB-1992; US-841646.
PR 01-NOV-1993; US-147023.
PA (STRYK) STRYKER CORP.
PI Kuberzampath T, Oppermann H, Ozkaynak E, Rueger DC;
DR WPI: 96-010159/01.
DR P-PSDB: R85757.
PT Antibodies with osteogenic protein binding specificity - used in
P1 purification of osteogenic proteins, and as antigenic proteins
PS Disclosure: Column 71-86; 129pp; English.
CC This sequence represents the genomic sequence which encodes the human
CC osteogenic protein, hOP-1. hOP-1 has homology with proteins in the
CC TGF-beta superfamily. Fragments of the hOP-1 protein can be used in
CC the production of dimeric peptides which may be used in the generation
CC of antibodies with binding specificities for osteogenic proteins. The
CC antibodies are capable of binding specifically to an epitope of the
CC osteogenic protein and may be used in purification protocols. Osteogenic
CC proteins, such as hOP-1, may be used in an implantable osteogenic device
CC which allows predictable bone formation to correct acquired and
CC congenital craniofacial and other skeletal or dental anomalies. They may
CC be used to induce local endochondral bone formation in non-union
CC fractures and in other clinical applications including dental and
CC periodontal applications where bone formation is required. Other
CC potential applications include cartilage repair, e.g. in the treatment
CC of osteoarthritis.
SQ Sequence 17410 BP; 3831 A; 4576 C; 4716 G; 4287 T;
Query Match 77.9%; Score 14.8; DB 1; Length 17410;
Best Local Similarity 88.9%; Pred. No. 94;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAGGATCAGACATGCC 18
DB 1312 AAGGAGCAGAAAAATGCC 1295
RESULT 9
ID T58758 standard; DNA: 3931 BP.
AC T58758
DE Murine long form CSF-1 coding sequence.
KW Long form; short form; human colony stimulating factor-1; HuCSF-1;
KW truncated protein; immunosuppression; chemotherapy;
KW bone marrow transplantation; AIDS; ss.
OS Mus musculus.
FH Key
FH Location/Qualifiers
FT cds
FT 160. 1818
FT /tag- a
FT /Product= Full_length_long_form_mucsf-1
FT signal_peptide 160. 225 ;

FT /tag- b
FT mat_peptide 256. 1815
FT /tag- c
FT US5573930-A.
PN 12-NOV-1996.
PD 05-FEB-1985; 698359.
PF 05-FEB-1985; US-698359.
PR 30-APR-1985; US-728834.
PR 14-JUN-1985; US-744924.
PR 18-JUL-1985; US-756814.
PR 21-JAN-1986; US-821068.
PR 20-JUN-1986; US-876819.
PR 24-OCT-1986; US-923067.
PR 16-APR-1987; US-939657.
PR 16-APR-1987; US-939654.
PR 27-NOV-1991; US-799411.
PR 27-NOV-1991; US-799039.
PR 28-DEC-1992; US-999298.
PA (CETUS) CETUS ONCOLOGY CORP.
PI Coyne MY, Halenbeck RF, Kawasaki ES, Kolts RE, Ladner MB;
PI Martin GA, Noble JA;
DR WPI: 96-517883/51.
DR P-PSDB: W10070.
PT DNA encoding human colony-stimulating factor-1 N-terminal deletion
PT mutants - opt. with C-terminal truncation(s), useful for treating
PT immunosuppression caused by e.g. chemotherapy or bone marrow
PT transplants
PS Disclosure: Column 61-68; 45pp; English.
CC The sequences given in T58758-59 encode the long and short forms of
CC murine colony stimulating factor (mucsf)-1. These sequences are closely
CC related to human CSF-1 sequences which were used in the design of the
CC truncated proteins of the invention (see also W10072-86). These
CC truncated proteins have 3 amino acids deleted at the N-terminal end,
CC and a variable number deleted from the C-terminal end. The truncated
CC proteins have mol. wts. much closer to those found in naturally
CC occurring CSF-1 dimers, and it is thought that natural CSF-1 may be
CC C-terminally truncated. The novel proteins may be useful for overcoming
CC immunosuppression induced e.g. by chemotherapy, bone marrow
CC transplantation or diseases, e.g. AIDS.
SQ Sequence 3931 BP; 973 A; 1106 C; 1004 G; 848 T;
Query Match 77.9%; Score 14.8; DB 1; Length 3931;
Best Local Similarity 88.9%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAGATCAGAACATGCCT 19
DB 542 AAGAGCAGAACAGGCCT 559
RESULT 10
ID T58759 standard; DNA: 1987 BP.
AC T58759;
DE Murine short form CSF-1 coding sequence.
KW Long form; short form; human colony stimulating factor-1; HuCSF-1;
KW truncated protein; immunosuppression; chemotherapy;
KW bone marrow transplantation; AIDS; ss.
OS Mus musculus.
FH Key
FH Location/Qualifiers
FT cds
FT 70. 1728
FT /tag- a
FT /Product= Full_length_short_form_mucsf-1
FT signal_peptide 70. 165
FT /tag- b
FT mat_peptide 166. 1725
FT /tag- c
FT US5573930-A.
PN 12-NOV-1996.
PF 05-FEB-1985; 698359.
PR 05-FEB-1985; US-698359.

CC N1A proteinase RNA (T47074) isolated from maize dwarf mosaic virus
CC (MDMV) (see also T47073) was amplified by PCR and modified to
CC include an Arg initiation codon and premature stop codon. The
CC construct was inserted into pCIB4421, contg. the PEP promoter and
CC CMV 35S terminator to give pCIB5018. The plasmid was used to
CC transform maize line CG00526 embryos. Plants that developed from
CC the embryos were screened for the transgene by PCR (see also
CC T47075-80). Seeds from selected lines were grown to plants and
CC these exposed to MDMV. Some plants were resistant, i.e. they
CC contained no detectable N1A transcripts and no virus.
SQ Sequence 1296 BP; 450 A; 252 C; 276 G; 2 T; 316 U;

Query Match 77.9%; Score 14.8; DB 1; Length 1296;
Best Local Similarity 72.2%; Pred. No. 69;
Matches 13; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 2 AGATCAGACATGCT 19
|||||
Db 78 AGAUCGAGAGAUCCU 95

RESULT 13
ID T80599 standard; cDNA; 3931 BP.
AC T80599; (first entry)
DE Murine 4 kb colony stimulating factor-1 encoding cDNA.
KW CSF-1; macrophage colony stimulating factor; M-CSF; immunostimulant;
KW immunosuppression; chemotherapy; transplant; AIDS; neoplasm;
KW acquired immune deficiency syndrome; antitumour; lymphokines;
KW mouse; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT signal_peptide 160..255
FT /*tag- a
FT mat_peptide 256..1816
FT /*tag- b
FT /product- CSF-1
PN US5643563-A.
PD 01-JUL-1997.
PE 05-FEB-1985; 698359.
PF 28-DEC-1992; US-999280.
PR 05-FEB-1985; US-698359.
PR 30-APR-1985; US-728834.
PR 14-JUN-1985; US-744924.
PR 18-JUL-1985; US-756814.
PR 21-JAN-1986; US-821068.
PR 20-JUN-1986; US-876819.
PR 24-OCT-1986; US-923067.
PR 16-APR-1987; US-039654.
PR 16-APR-1987; US-039657.
PR 13-OCT-1987; US-105261.
PR 27-NOV-1991; US-799039.
PR 27-NOV-1991; US-799411.
PR 21-APR-1995; US-426036.
PA (CHIR) CHIRON CORP
PI Coyne MY, Halebek RF, Kawasaki ES, Kohts KE, Ladner MB;
PI Martin GA, Noble JA;
PI WPI: 97-350183/32.
DR P-PSDB: W22615.
PT Carboxy-truncated colony-stimulating factor 1 polypeptide(s) - as
PT immunostimulants for treating immunosuppression
PS Disclosure: Column 63-70; 45pp; English.
CC The present sequence encodes a novel murine 4 kb colony stimulating
CC factor (CSF-1). Human CSF-1 (also known as macrophage colony stimulating
CC factor, M-CSF) proteins are useful as immunostimulants for treating
CC immunosuppressed patients, e.g. chemotherapy patients, transplant
CC recipients or AIDS (acquired immune deficiency syndrome) patients,
CC or for treating neoplasms and infections. They may be administered
CC together with antitumour agents or lymphokines.
SQ Sequence 3931 BP; 973 A; 1106 C; 1004 G; 848 T;

Query Match 77.9%; Score 14.8; DB 1; Length 3931;
Best Local Similarity 88.9%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 2 AGATCAGACATGCT 19
|||||
Db 542 AGACAGACAGAGGCT 559

RESULT 14
ID T80600 standard; cDNA; 1987 BP.
AC T80600; (first entry)
DE Murine 2 kb colony stimulating factor-1 encoding cDNA.
KW CSF-1; macrophage colony stimulating factor; M-CSF; immunostimulant;
KW immunosuppression; chemotherapy; transplant; AIDS; neoplasm;
KW acquired immune deficiency syndrome; antitumour; lymphokines;
KW mouse; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT signal_peptide 70..165
FT /*tag- a
FT mat_peptide 166..1726
FT /*tag- b
FT /product- CSF-1
PN US5643563-A.
PD 01-JUL-1997.
PE 05-FEB-1985; 698359.
PF 28-DEC-1992; US-999280.
PR 05-FEB-1985; US-698359.
PR 30-APR-1985; US-728834.
PR 14-JUN-1985; US-744924.
PR 18-JUL-1985; US-756814.
PR 21-JAN-1986; US-821068.
PR 20-JUN-1986; US-876819.
PR 24-OCT-1986; US-923067.
PR 16-APR-1987; US-039654.
PR 16-APR-1987; US-039657.
PR 13-OCT-1987; US-105261.
PR 27-NOV-1991; US-799039.
PR 27-NOV-1991; US-799411.
PR 21-APR-1995; US-426036.
PA (CHIR) CHIRON CORP
PI Coyne MY, Halebek RF, Kawasaki ES, Kohts KE, Ladner MB;
PI Martin GA, Noble JA;
PI WPI: 97-350183/32.
DR P-PSDB: W22616.
PT Carboxy-truncated colony-stimulating factor 1 polypeptide(s) - as
PT immunostimulants for treating immunosuppression
PS Disclosure: Column 71-76; 45pp; English.
CC The present sequence encodes a novel murine 2 kb colony stimulating
CC factor (CSF-1). Human CSF-1 (also known as macrophage colony stimulating
CC factor, M-CSF) proteins are useful as immunostimulants for treating
CC immunosuppressed patients, e.g. chemotherapy patients, transplant
CC recipients or AIDS (acquired immune deficiency syndrome) patients,
CC or for treating neoplasms and infections. They may be administered
CC together with antitumour agents or lymphokines.
SQ Sequence 1987 BP; 496 A; 569 C; 511 G; 411 T;

Query Match 77.9%; Score 14.8; DB 1; Length 1987;
Best Local Similarity 88.9%; Pred. No. 73;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 2 AGATCAGACATGCT 19
|||||
Db 452 AGACAGACAGAGGCT 469

RESULT 15
ID T94752; (first entry)

```

ID T94752, standard: cDNA, 3931 BP.
AC T94752;
DT 18-FEB-1998 (first entry)
DE Murine colony stimulating factor 1 long form encoding cDNA.
KW Human: colony stimulating factor 1; CSF-1; truncation; chemotherapy;
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 160..1818
FT FT /tag- a
FT FT /product- Colony_stimulating_factor_1
FT FT /note- "long form"
FT sig_peptide 160..254
FT mat_peptide 255..1815
FT /tag- b
FT /tag- c
PN US5681719-A.
PD 28-OCT-1997.
PF 05-FEB-1985; 698359.
PR 28-DEC-1992; US-9989280.
PR 05-FEB-1985; US-698359.
PR 30-APR-1985; US-728834.
PR 14-JUN-1985; US-744924.
PR 18-JUL-1985; US-756814.
PR 21-JAN-1986; US-821068.
PR 20-JUN-1986; US-876819.
PR 24-OCT-1986; US-923067.
PR 16-APR-1987; US-039654.
PR 16-APR-1987; US-039657.
PR 13-OCT-1987; US-105261.
PR 27-NOV-1991; US-799039.
PR 27-NOV-1991; US-799411.
PR 08-MAR-1995; US-401013.
PA (CHIR-) CHIRON CORPORATION.
PI Coyne MY, Halenbeck RF, Kawasaki ES, Roths KE, Ladner MB;
PI Martin GA, Noble JA;
DR WPI: 97-535051/49.
DR P-PSDB: W35752.
PT DNA encoding amino-truncated colony stimulating factor 1 - useful as
PT immunostimulant in treatment of chemotherapy patients
PS Disclosure; Column 61-66; 45pp; English.
CC The present sequence encodes the long form of mouse colony stimulating
CC factor 1 (CSF-1). The present invention has developed DNA molecules
CC which encode amino truncated forms of CSF-1 e.g. specifically
CC claimed DNA which encodes a human CSF-1 polypeptide comprising amino
CC acids 4-221 of the 536 amino acid sequence W35750 encoded by T94747;
CC where dimers of the polypeptide have CSF-1 activity. The products can
CC be used for the recombinant production of amino-terminal and/or
CC carboxy-terminal truncated CSF-1, which is useful as an immunostimulant
CC in the treatment of chemotherapy patients, e.g. bone marrow graft
CC recipients. Compared to genes encoding so called mature forms preceded
CC by a methionine start codon, genes encoding amino-terminal deleted
CC forms in bacteria give rise to expression products which are more
CC efficiently processed to remove the amino-terminal methionine, and
CC which are comparatively homogenous, as judged by reverse phase HPLC.
SO Sequence 3931 BP; 973 A; 1106 C; 1004 G; 848 T;

```

```

Query Match 77.9%; Score 14.8; DB 1; Length 3931;
Best Local Similarity 88.9%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 AGATGAGAACATGGCT 19
   ||||| ||||| |||||
DB 542 AGGAGCGAGAACAGGCT 559

```

Search completed: September 13, 1999, 15:55:35
 Job time: 4802 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:56 ; Search time 1694.61 Seconds
(without alignments)
22.116 Million cell updates/sec

Title: US-09-325-095-24

Perfect score: 19
Sequence: 1 AAGCATCAGACATGCTT 19

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:
1: em_est1.*
2: em_est2.*
3: em_est3.*
4: em_est4.*
5: em_est5.*
6: em_est6.*
7: em_est7.*
8: em_est8.*
9: em_est9.*
10: em_est10.*
11: em_est11.*
12: em_est12.*
13: em_est13.*
14: em_est14.*
15: em_est15.*
16: em_est16.*
17: em_est17.*
18: em_est18.*
19: em_est19.*
20: em_est20.*
21: em_est21.*
22: em_est22.*
23: em_est23.*
24: em_est24.*
25: em_est25.*
26: em_est26.*
27: em_est27.*
28: em_est28.*
29: em_est29.*
30: em_est30.*
31: em_est31.*
32: em_est32.*
33: em_est33.*
34: em_est34.*
35: em_est35.*
36: em_est36.*
37: em_est37.*
38: em_est38.*
39: em_est39.*
40: em_est40.*
41: em_est41.*
42: em_est42.*
43: em_est43.*
44: em_est44.*
45: em_est45.*
46: em_est46.*
47: em_est47.*
48: em_est48.*
49: em_est49.*
50: em_est50.*
51: em_est51.*
52: em_est52.*
53: em_est53.*

54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17.4	91.6	417	29	AA190409	AA190409 zp89c10.r
2	16.4	86.3	479	37	AA731088	AA731088 mw88a08.s
3	16.4	86.3	485	38	AA767230	AA767230 n280h01.s
4	16.4	86.3	918	38	AA767233	AA767233 n280h05.s
5	16.4	86.3	561	41	AI017576	AI017576 ou23f12.x
6	16.4	84.2	634	29	AA151764	AA151764 zo3f06.s
7	16.4	84.2	522	35	AA581298	AA581298 n43e06.s
8	16.4	84.2	509	47	AI475209	AI475209 t180b03.x
9	16.4	84.2	468	47	AI499908	AI499908 t003f05.x
10	16.4	84.2	414	49	AI640828	AI640828 t269a10.x
11	15.8	83.2	651	29	AA142034	AA142034 CR00149.3
12	15.8	83.2	406	29	AA146711	AA146711 zo35e10.r
13	15.8	83.2	235	31	AA288447	AA288447 vb37a12.r
14	15.8	83.2	403	35	AA593545	AA593545 nm28c07.s
15	15.8	83.2	578	37	AA672725	AA672725 vp07d01.r
16	15.8	83.2	488	38	AA759578	AA759578 vv55c04.r
17	15.8	83.2	260	41	AI048755	AI048755 ud32a10.r
18	15.8	83.2	638	47	AI444697	AI444697 486015c11
19	15.8	83.2	276	47	AI462631	AI462631 vb37a12.x
20	15.8	83.2	641	47	AI467712	AI467712 486015c11
21	15.8	83.2	468	48	AI596476	AI596476 v174h10.x
22	15.8	83.2	491	54	HS0010092	AI045242 Homo sapi
23	15.4	81.1	504	38	AA760081	AA760081 vv69c06.r
24	15.4	81.1	500	39	AA898726	AA898726 NCM6E107
25	15.4	81.1	509	42	AI142259	AI142259 q962d07.r
26	15.4	81.1	294	29	AA175134	AA175134 ms82g07.r
27	15.4	78.9	709	41	AI054610	AI054610 csm00011
28	15.4	78.9	372	46	AI448592	AI448592 mc50b09.x
29	15.4	78.9	370	46	AI450814	AI450814 ms82g07.x
30	15.4	78.9	370	49	AI639779	AI639779 ms82g07.y
31	14.8	77.9	218	20	220318	220318 HSAABXHU P
32	14.8	77.9	286	20	221098	221098 HSAABXMX T
33	14.8	77.9	320	21	F12657	F12657 HSC3G061 D
34	14.8	77.9	429	21	T74316	T74316 YC84d06.r1
35	14.8	77.9	358	21	T78190	T78190 Yd79d12.r1
36	14.8	77.9	479	24	H75125	H75125 250 P1RFG1
37	14.8	77.9	304	26	W20869	W20869 mb96c04.r1
38	14.8	77.9	488	26	W45995	W45995 mc81g05.r1
39	14.8	77.9	486	27	AA021817	AA021817 mb84e01.r
40	14.8	77.9	481	29	AA187027	AA187027 zp72g01.r
41	14.8	77.9	380	30	AA240477	AA240477 mw34b01.r
42	14.8	77.9	340	30	AA247156	AA247156 chp035.s
43	14.8	77.9	482	30	AA259669	AA259669 va54d09.r
44	14.8	77.9	440	30	AA260212	AA260212 va88f07.r
45	14.8	77.9	652	53	HS007439	AI042589 Homo sapi

ALIGNMENTS

RESULT 1
AA190409
LOCUS
DEFINITION
IMAGE:627378 5' similar to SW:P11A.BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
mRNA sequence.
AA190409
ACCESSION
AA190409
417 bp
mRNA
EST
15-JAN-1997
zp89c10.r1 Stralagene Hela cell s3 937216 Homo sapiens cDNA clone
IMAGE:627378 5' similar to SW:P11A.BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
mRNA sequence.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2150988.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert length: 1003 Std Error: 0.00
Seq primer: -40ml3 fwd: ET from Amersham.

FEATURES
source Location/Qualifiers

1..485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21q"
/clone="IMAGE:1301809"
/clone_lib="NCI-CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

/note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCACTGTGAGTGGAGCGCCGCTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 87 a 125 c 107 g 166 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 38; Length 485;
Best Local Similarity 94.4%; Pred. NO.1.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AGGATCAGAACATGCCT 19
||||| |||||||||
DB 435 AGGATAGAACATGCCT 418

RESULT 4
AA767233 918 bp mRNA EST 08-FEB-1998
LOCUS n280h05.s1 NCI CGAP GCB1 Homo sapiens cDNA clone IMAGE:1301817 3'
DEFINITION similar to TR:Q92923 Q92923 SWI/SNF COMPLEX 170 KDA SUBUNIT. ;,
mRNA sequence.

ACCESSION AA767233
NID 92818248
VERSION AA767233.1 GI:2818248
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 918)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2150992.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert length: 1580 Std Error: 0.00
Seq primer: -40ml3 fwd: ET from Amersham
High quality sequence stop: 484.

FEATURES
source Location/Qualifiers

1..918
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11; 21q"
/clone="IMAGE:1301817"
/clone_lib="NCI-CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

/note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCACTGTGAGTGGAGCGCCGCTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 194 a 220 c 230 g 274 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 38; Length 918;
Best Local Similarity 94.4%; Pred. NO.1.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AGGATCAGAACATGCCT 19
||||| |||||||||
DB 435 AGGATAGAACATGCCT 418

RESULT 5
A1017576 561 bp mRNA EST 27-AUG-1998
LOCUS ou23ef2.x1 Soares NFL_T GBC S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1627151.3' similar to TR:Q92923 Q92923 SWI/SNF COMPLEX 170
KDA SUBUNIT. ;, mRNA sequence.

ACCESSION A1017576
NID 93231912
VERSION A1017576.1 GI:3231912
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 561)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

COMMENT On Jan 14, 1998 this sequence version replaced gl:1798155.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 988 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 477.
Location/Qualifiers

FEATURES

source

```
1. 561
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pRTD-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGP, GCBI) were mixed, and ss circles were made in
vitro. Following BAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682633-687239,
726408-728711, and 729096-733399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
141 c 127 g 182 t
```

BASE COUNT

111 a

Query Match 86.3%; Score 16.4; DB 41; Length 561;
Best Local Similarity 94.4%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AGCATCAGAACATGCT 19

||||| |||||||

DB 428 AGGATAGAACATGCT 411

RESULT 6
AA151764/c 634 bp mRNA EST 06-AUG-1997
LOCUS
DEFINITION
z30306.s1 Striatogene colon (#937204) Homo sapiens cDNA clone
IMAGE:588419.3 similar to SW:RL3_HUMAN P39023 60S RIBOSOMAL
PROTEIN L3. ; mRNA sequence.

ACCESSION AA151764
NID 91720478
VERSION AA151764.1 GI:1720478
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 634)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapel, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Scheleberg, K., Soares, M.B., Tan, E., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.

TITLE
JOURNAL
MEDLINE
COMMENT
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 688 Std Error: 0.00
Seq primer: -40m13 fwd. from Amersham
High quality sequence stop: 391.
Location/Qualifiers

FEATURES

source

```
1. 634
/organism="Homo sapiens"
/db_xref="GDB:4620721"
/db_xref="taxon:9606"
/clone_lib="IMAGE:588419"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: colon; Vector: pBluescript SK-; Site_1:  
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:  
01190 dt. T-84 colonic epithelial cell line. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' GATTGGGACGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTCTTTTCTTTTCTTTT 3'"
137 a 169 c 156 g 163 t 9 others
```

BASE COUNT

137 a

Query Match 84.2%; Score 16; DB 29; Length 634;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GATCAGACATGCT 19

||||| |||||||

DB 371 GATCAGACATGCT 356

RESULT 7
AA581298 522 bp mRNA EST 05-JAN-1998
LOCUS
DEFINITION
nd43e06.s1 NCI-CGAP_R1 Homo sapiens cDNA clone IMAGE:803074.3;
sequence.
similar to SW:RL3_HUMAN P39023 60S RIBOSOMAL PROTEIN L3. ; mRNA
sequence.

ACCESSION AA581298
NID 92359070
VERSION AA581298.1 GI:2359070
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 522)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
COMMENT
On May 9, 1995 this sequence version replaced gl:804204.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Lee Helman, Ph.D., Michael Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: Life Technologies Inc., David Krizman,
Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -4m13 fwd. ET from Amersham
High quality sequence stop: 321.
Location/Qualifiers
1. 522
/organism="Homo sapiens"

FEATURES

source

/db_xref="taxon:9606"
/clone_image="803074"
/clone_id="NCI_CGAP_A1"
/tissue_type="bulk alveolar tumor"
/lab_host="DH10B"
/note="Vector: PCMV-SPOPT2; Site_1: SalI; Site_2: NotI;
Cloned unidirectionally. Primer: Oligo dt."

BASE COUNT 114 a 138 c 126 g 144 t
ORIGIN

Query Match 84.2%; Score 16; DB 35; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GATCAGAACATGCCT 19
|||||
Db 379 GATCAGAACATGCCT 364

RESULT 8
A1475209/c 509 bp mRNA EST 14-APR-1999
LOCUS t180b03.x1 NCI CGAP Col4 Homo sapiens cDNA clone IMAGE:2153357 3'
DEFINITION similar to SW:R13_HUMAN P39023 60S RIBOSOMAL PROTEIN L3. ; mRNA
sequence.
ACCESSION A1475209
NID 94328254
VERSION A1475209.1 GI:4328254
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 509)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246645.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert Length: 1472 Std Error: 0.00
Seq primer: -40UP from G1bco
High quality sequence stop: 316.
Location/Qualifiers
1. 509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2153357"
/clone_id="NCI_CGAP_Col4"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: PCMV-SPOPT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.7 kb. Life Technologies catalog #:
11531-019"

BASE COUNT 111 a 135 c 119 g 144 t
ORIGIN

Query Match

84.2%; Score 16; DB 47; Length 509;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GATCAGAACATGCCT 19
|||||
Db 383 GATCAGAACATGCCT 368

RESULT 9
A1499908/c 468 bp mRNA EST 12-MAY-1999
LOCUS t003f05.x1 NCI CGAP ut2 Homo sapiens cDNA clone IMAGE:2177985 3'
DEFINITION similar to SW:R13_HUMAN P39023 60S RIBOSOMAL PROTEIN L3. ; mRNA
sequence.
ACCESSION A1499908
NID 94391890
VERSION A1499908.1 GI:4391890
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 468)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 20, 1998 this sequence version replaced gi:2979745.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert Length: 1395 Std Error: 0.00
Seq primer: -40UP from G1bco
High quality sequence stop: 338
POLYA-No.
Location/Qualifiers
1. 468
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2177985"
/clone_id="NCI_CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: PCMV-SPOPT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

BASE COUNT 104 a 120 c 99 g 145 t
ORIGIN

Query Match 84.2%; Score 16; DB 47; Length 468;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GATCAGAACATGCCT 19
|||||
Db 389 GATCAGAACATGCCT 374

RESULT 10
A1640828/c

LOCUS	A1640828	414 bp	mRNA	EST	28-APR-1998
DEFINITION	t269a10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2293818 3' similar to SW:R13_HUMAN P39023 60S RIBOSOMAL PROTEIN L3 /, mRNA sequence.				
ACCESSION	A1640828				
NID	g4703937				
VERSION	A1640828.1	GI:4703937			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheta; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 414)				
AUTHORS	*NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	On May 18, 1998 this sequence version replaced gi:3136952.				
	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Life Technologies catalog #: 11548-013 DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/JMLM at: www-bio.liml.gov/dbrrp/image/image.html				
FEATURES	Seq primer: -40UP from GIBCO High quality sequence stop: 337. Location/Qualifiers				
SOURCE	1..414 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2293818" /clone_lib="NCI_CGAP_Pan1" /tissue_type="adenocarcinoma" /lab_host="DH10B" /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: oligo dT" Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"				
BASE COUNT	93 a 112 c 94 g 115 t				
ORIGIN					
	Query Match 84.28; Score 16; DB 49; Length 414; Best Local Similarity 100.08; Prid. No. 1.9e+02; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	4 GATCAGACATGCGCT 19 				
TH	374 GATCAGACATGCGCT 359				
RESULT 11					
LOCUS	A1142034	651 bp	mRNA	EST	29-NOV-1998
DEFINITION	CK00149.3prine CK Drosophila melanogaster embryo Bluescript				
LOCUS	Drosophila melanogaster cDNA clone CK00149 3prime, mRNA sequence.				
ACCESSION	A1142034				
NID	g1703931				
VERSION	A1142034.1	GI:1703931			
KEYWORDS	EST.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Endopterygota; Diptera; Brachyera;				
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
	1 (bases 1 to 651)				
REFERENCE	Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,				
AUTHORS	Brokstein,P., Lewis,S. and Rubin,G.M.				
TITLE	BDGP/HMI Drosophila EST project				

JOURNAL
 COMMENT
 Unpublished (1997)
 On Jun 18, 1996 this sequence version replaced gi:1366617.
 Other ESTs: CK00149..sprime
 Contact: Harvey, D.
 G. M. Rubin-Molecular and Cell Biology
 University of California Berkeley
 539 LSA, Berkeley, CA 94720-3200, USA
 Fax: 510 643 9947
 Email: <http://www.fruitfly.org/EST/estfruitfly.berkeley.edu>
 Insert Length: 651 Std Error: 0.00
 Plate: 2 row: C column: 2 651.
 High quality sequence stop: 651.
 Location/Qualifiers
 1..651
 /organism="Drosophila melanogaster"
 /db_xref="BDGP_EST:BDCL000120"
 /db_xref="taxon:7227"
 /clone="CK00149"
 /clone_1lb="CK Drosophila melanogaster embryo Bluescript"
 /sex="male and female"
 /dev_stage="0 to 24 hours old embryo"
 /lab_host="XLI Blue MRF"
 /note="Organ: embryo; Vector: Bluescript SK; Site_1: ClaI
 Site_2: PstI; mRNA purified from rough endoplasmic
 reticulum-bound polysomes was used as a template. cDNA's
 directionally cloned at HindIII and PstI in Bluescript
 SK+. Cloned into HindIII, but the site was destroyed to
 add an adapter sequence. So ClaI is now the restriction
 enzyme at site 1 of vector. Primers - 5' universal, 3'
 m13-20 (reverse), 3' T3, 5' T7." 5 others
 m13-20 (reverse), 3' T3, 5' T7." 5 others

BASE COUNT
 165 a 155 c 139 g 187 t

ORIGIN

Query Match 83.2% Score 15.8: DB 29: Length 651:
 Best Local Similarity 89.5% Pred. NO. 2.4e-02;
 Matches 17: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

Oy 1 AAGATCAGACCAATGCCT 19
 ||| |||||
 Db 585 AAGACTCAGACCAATGCCT 567

RESULT 12
 AA146711 406 bp mRNA EST 05-DEC-1996
 LOCUS
 DEFINITION zc35e10.1 StrataGene colon (#937204) Homo sapiens cDNA clone
 IMAGE:588906 5' similar to SW:STAL.HUMAN P42224 SIGNAL TRANSDUCER
 AND ACTIVATOR OF TRANSCRIPTION 1-ALPHA/BETA ; mRNA sequence.
 ACCESSION AA146711
 NID g1716152
 VERSION AA146711.1 GI:1716152
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 406)
 Hallier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapellin,B.,
 Chissoc,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
 Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
 Mavdis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
 Rollfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thiermy-Meg,J.,
 Trevisakis,E., Underwood,K., Wohlmann,P., Waterson,R., Wilson,R.
 and Maria,M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 On Sep 12, 1996 this sequence version replaced gi:1393797.

TITLE
 JOURNAL
 MEDLINE
 COMMENT

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 255.
 Location/Qualifiers

FEATURES

Source

1. 406
 /organism="Homo sapiens"
 /db_xref="GDB:4626646"
 /db_xref="taxon:9606"
 /clone="IMAGE:588906"
 /clone_lib="Scratchgene colon (#937204)"
 /lab_host="SOLR cells (Kanamycin resistant)"
 /note="Organ: colon; Vector: pBluescript SK-; Site: 1;
 EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. T-84 colonic epithelial cell line. Average
 insert size: 1.0 kb; unl-zap xr vector: -5' adaptor
 sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT 122 a 83 c 102 g 98 t 1 others
 ORIGIN

Query Match 83.2%; Score 15.8; DB 29; Length 406;
 Best Local Similarity 89.5%; Pred. No. 2.4e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAGGATCAGACATGCCT 19
 ||||| ||||| |||||
 Db 318 AAGGACCGAATAATGCCT 336

RESULT 13
 AA288447 235 bp mRNA EST 14-APR-1997
 LOCUS VB37A12.1 Soares mouse lymph node NBM1N Mus musculus cDNA clone
 DEFINITION IMAGE:751102 5', mRNA sequence.
 ACCESSION AA288447
 NID 91937377
 VERSION AA288447.1 GI:1937377
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 REFERENCE 1 (bases 1 to 235)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucada, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 TITLE The WashU-HM1 Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On May 17, 1993 this sequence version replaced gi:426136.

CONTACT: Marra M/Mouse EST Project
 WashU-HM1 Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:460086
 Putative full length read
 vector to vector length is 239
 Seq primer: -28M13 rev2 ET from Amersham.

FEATURES

Source

1. 235
 /organism="Mus musculus"

/strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:751102"
 /clone_lib="Soares mouse lymph node NBM1N"
 /sex="male"
 /tissue="type="lymph node"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: lymph node; Vector: pTTT3D-Pac (Pharmacia)
 with a modified polylinker; Site 1: Not I; Site 2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 [5',
 TGTACCAATGTGAAGTGGAGCGCCGCCGACATCTTTTCTTTTCTTTTCTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pTTT3 vector. RNA
 provided by Dr. Bertrand Jordan. Library constructed and
 normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 85 a 48 c 49 g 53 t
 ORIGIN

Query Match 83.2%; Score 15.8; DB 31; Length 235;
 Best Local Similarity 89.5%; Pred. No. 2.4e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAGGATCAGACATGCCT 19
 ||||| ||||| |||||
 Db 151 AAGGATCGCAACATGCCT 169

RESULT 14
 AA593545 403 bp mRNA EST 25-SEP-1997
 LOCUS nm28c07.s1 NCI CGAP Gas1 Homo sapiens cDNA clone IMAGE:1085196 3'
 DEFINITION similar to contains element MBR8 repetitive element; mRNA
 sequence.
 ACCESSION AA593545
 NID 92409307
 VERSION AA593545.1 GI:2409307
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 403)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 9, 1995 this sequence version replaced gi:802451.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnl.gov/bhrp/image/image.html

Insert length: 956 Std Error: 0.00
 Seq primer: -40M13 fwd. ET from Amersham
 High quality sequence stop: 380.
 Location/Qualifiers

FEATURES

Source

1. 403
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="12p"
 /clone="IMAGE:1085196"

```

/clone_lib="NCI_GAP_Gast1"
/issue_type="gastric tumor"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: stomach; Vector: Bluescript SK-; Site: 1;
EcORI: Site 2: XhoI; Cloned unidirectionally. Primer:
5' GATTGGCAGCAG 3' 3' adaptor sequence: 5'
CTCGAGTGTGTGTGTGTGT 3' Average insert size: 1.0 kb."

BASE COUNT      112 a      92 c      95 g      104 t
ORIGIN

Query Match      83.2%; Score 15.8; DB 35; Length 403;
Best Local Similarity 89.5%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAGGATCAGAACATGCT 19
      ||| | ||| ||| ||| |||
Db      12 AAGCACCAGAACATGCT 30

RESULT 15      AA672725      578 bp      mRNA      EST      25-NOV-1997
LOCUS      AA672725/c      VP07601.F1 Soares mouse mammary gland NBMWG Mus musculus cDNA clone
DEFINITION      IMAGE:1067905 5', mRNA sequence.
ACCESSION      AA672725
NID      92644942
VERSION      AA672725.1 GI:2644942
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 578)
Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Schellberg, K., Sepcoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 19, 1997 this sequence version replaced gi:1520426.

TITLE      The WashU-HMI Mouse EST Project
JOURNAL
COMMENT

Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:590265
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 498.
Location/Qualifiers
1. 578
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1067905"
/clone_lib="Soares mouse mammary gland NBMWG"
/sex="male"
/issue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: p773D-Pac
(Pharmacia) with a modified polylinker; Site: 1;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer 15'
TGTACCATCTGAGTGGAGCGCGGGAATGTTTGTGTGTGTGTGTGTGT
T 3'; double-stranded cDNA was ligated to Eco RI

```

```

adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT      142 a      121 c      137 g      178 t
ORIGIN

Query Match      83.2%; Score 15.8; DB 37; Length 578;
Best Local Similarity 89.5%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAGGATCAGAACATGCT 19
      ||| | ||| ||| ||| |||
Db      148 AAGGTCAGAACATGACT 130

Search completed: September 13, 1999, 15:45:56
Job time: 4044 sec

```


CDS

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBAR-1"
1..3207
/EC_number="2.7.1.67"
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="AAB62534.1"
/db_xref="pid:92245506"
/db_xref="GI:2245506"
/translation="MPPRSSGELMGILHMPRIIVECLIPNGMIVTECLREATLT
IKHEFKARKRYPLVQLQDSESYIFVYTOGEAREEEFEDTRLCIDLQDPEIKVI
EYGNREERILNREIGFAGIMPCEDVADPEVDERRNINLVCKEAVDANAPH
SRMAYIPPNVSSPELPHNINLNDKQIIVIVIVISPNNDKQYTLKINHCVPE
OYIAEIRKTRSMILSSEQLKCYLEYGKYLKVCGDEFLKPLSQYKIRSC
IMLGMPNIMLAKESLTOLPLDTFMPSSRSISTATPYNGCATKSLMTINSAL
RIRICATVYVNIIRDIKIVRTGIYHGEPLCDNNTORVPCSNPRNEMLVYDI
IPDLPRARALCISTSVYGRKGAKEHCPLMGNINMEDIYDTLVSGMALNLAVPH
GLEDLNPIGVYGSNPNKTEPCLELEFDWFSNPVPEPMVYIEEHANWTISREGENY
SHAGLSNRILARDNELRESDKQLRAICTRDPSEITTEDEKDFLMSHRHCVTIPILP
KLLSVKNSRDEVAQMYCLVADMPPIKEQAMELDCNYPDPVRAFAVCLKELYT
DKLSQYLQIVLYKYOVLQYLDNOLVRELKALNORIGHEFEFHLKSEHNTKVSQ
RGILLSEYCRACGAVLKHLISROYAMKELINTDILKOEKDETOKYOMKEIYBOMR
RPFMDALQGFISPLNPAHQGNLBLECRIMSSAKRPLMNMENPDIMSELTFONE
IIFKNGDDLRODMTLQITRIMENIMWQGNLDRLMPLGCLSIGCVGLIEVYRNSHT
IMQIOCKGGLKQALQFNSHTLHOMKDNKGEITDPAIDLFTRSCAGVATFIIIG
DHRNSNIWKGQALQFHDGFHIDFKKRGYKREVPFVLQDPLVISGADQECT
KTRFERPOEMCKYALAIROHANLEINLFPMGLSGMPELQSFODIAYIRKTLADK
TEQEALEYFMKONDAHNGWTTKMDWIFHTIKOHALN"

```

BASE COUNT 1028 a 581 c 680 g 918 t

Query Match 100.0%; Score 19; DB 3; Length 3207;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCGTTCTTTAGGCATCA 19
 ||||||||||||||||
 Db 878 AGCGTTCTTTAGGCATCA 860

RESULT 2
 LOCUS AF001076 3452 bp mRNA VRT 08-JUL-1997
 DEFINITION Gallus gallus phosphoinositide 3-kinase catalytic subunit mRNA,
 complete cds.
 ACCESSION AF001076
 NID 92245505
 VERSION AF001076.1 GI:2245505
 KEYWORDS
 SOURCE chicken.
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
 Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 3452)
 AUTHORS Chang,H.W., Aoki,M., Fruman,D., Auger,K.R., Bellacosa,A.,
 Tschilis,P.N., Cantley,L.C., Roberts,T.M. and Vogt,P.K.
 TITLE Transformation of chicken cells by the gene encoding the catalytic
 subunit of PI 3-kinase
 JOURNAL Science 276 (5320), 1848-1850 (1997)
 MEDLINE 97334438
 REFERENCE 2 (bases 1 to 3452)
 AUTHORS Chang,H.W., Aoki,M. and Vogt,P.K.
 TITLE Direct Substitution
 JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
 Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
 BCC-239, La Jolla, CA 92037, USA
 FEATURES
 source
 1..3452
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /tissue_type="Brain"

5'UTR
 gene
 1..150
 151..3354
 /gene="c-p3k"
 /note="p3k proto-oncogene"
 151..3357
 /note="c-p3k"
 CDS
 151..3357

BASE COUNT 1073 a 672 c 783 g 924 t

Query Match 100.0%; Score 19; DB 4; Length 3452;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCGTTCTTTAGGCATCA 19
 ||||||||||||||||
 Db 1028 AGCGTTCTTTAGGCATCA 1010

RESULT 3
 LOCUS A37232 3498 bp DNA PAT 05-MAR-1997
 DEFINITION Sequence 1 from Patent WO9403609.
 ACCESSION A37232
 NID 92294345
 VERSION A37232.1 GI:2294345
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 3498)
 AUTHORS Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
 TITLE EUCHARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
 JOURNAL PATENT: WO 9403609-A 1 17-FEB-1994;
 IMP CANCER RES TRCH (GB)
 COMMENT Other publication JP 8503124T 960409.
 FEATURES
 source
 1..3498
 /organism="unidentified"
 /db_xref="taxon:32644"
 1..3207
 /note="unnamed protein product"

CDS
 1..3207
 /codon_start=1
 /protein_id="CAA02284.1"
 /db_xref="pid:e304808"
 /db_xref="PID:g2294346"
 /db_xref="GI:2294346"

/translation-"MPRPSGELMGHIMPRLIVECLPNGMIVTLECREATIT
IKHLEFEARKYPLHOLLODESSITFVSVOEARREPEFDETRLCRLFOPLKAT
EPVGNREKILNREIGFALGMPCEFDVNDPEVODFRNRLINVCRAVDRLDINSBH
SRMAYVPRPVSSEPELPHKYNKLDGQIIVYIVWVSPNNOKYTKLNNHCVE
QVIAEARKTRSMLSSEDLKLCVLEOYKYLKVCGEDEPLEKYSQYKYLISC
IMGRMNLMMAKESLSYLSOLPMDCFPMPSYSRISTATPMNGESTKSLWYNSL
RIKICATYVNLNRIIDIKYVRGITHGGEPLCDNYNTQVPCSNRMENLNDYI
IPDLPRARCLSLSCYKGRKGAKEEPLAMGINTLPDITDILVSGMAMLVNPFVH
GLEDLNPIDVGTSPNPKETPCLELEDFWSSVVKFPMVSEHNAWSREAGFESY
SHAGLSRLARNDELRENDKOLRAICTRDLSEITDEKDFLMSHRHAYCVTIPILP
KLIVSVKNSRDEVAQMYCLVKDMPPIKPEQAMELLDCNPDPMVRFGEAVCEKYL
DDKLSOYLIDLVYLYKYOYLNDNLVRFELKALTNORIGHFPMHLSKEMHNTVSO
RFGLLSYSCBACGMYLKHNRYEAKELINLIDIKORKEDETKYOKKFEVYEQNR
IPFNSGDLQODMLTIOITRIMENIMQNGDLMLPYGLSIGDCVGLLEVYNSHT
IMJOCKGKLGALQFNSHTLHOMLKONKEITDAIDLETSCACVATFVLIG
DRHNSNIMVDDQGLFHDGFHLDHKKRKFGRKREVPFLVODFLIVISKGOECT
KTRFEFEQECYKAYLAIRQHANLFTNLSMMLGSGMPELQSPDDIAYIRKTLALDK
TEOALVEPKOMNDAHGGWTTKMDIFRTHKOHALN"

BASE COUNT 1147 a 626 c 722 g 1003 t
ORIGIN

Query Match 100.0%; Score 19; DB 5; Length 3498;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGCTTCTTAGCATCA 19
|||||
Db 878 AGGCTTCTTAGCATCA 860

RESULT 4
HSPH13K 3424 bp mRNA PRI 24-AUG-1995
LOCUS H.sapiens mRNA for phosphatidylinositol 3-kinase.
DEFINITION 229090
ACCESSION 9472990
VERSION 229090.1 GI:472990
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1068)
AUTHORS Volinia,S., Hiles,I., Ormondroyd,E., Nizetic,D., Antonacci,R.,
Rocchi,M. and Waterfield,M.D.
TITLE Molecular cloning, cDNA sequence, and chromosomal localization of
the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene
JOURNAL Genomics 24 (3), 472-477 (1994)
MEDLINE 95229146
REFERENCE 2 (bases 1 to 3424)
AUTHORS Volinia,S.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig
Institute for Cancer, Research, 91 Riding House Street, London, W1P
8BT, UK

FEATURES
source
1..3424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="K562"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CAA82333.1"
/db_xref="PIR:P107291"
/db_xref="GI:472991"
/db_xref="SWISS-PROT:P42336"
/translation="MPRPSGELMGHIMPRLIVECLPNGMIVTLECREATIT
IKHLEFEARKYPLHOLLODESSITFVSVOEARREPEFDETRLCRLFOPLKAT
EPVGNREKILNREIGFALGMPCEFDVNDPEVODFRNRLINVCRAVDRLDINSBH

SRMAYVPRPVSSEPELPHKYNKLDGQIIVYIVWVSPNNOKYTKLNNHCVE
QVIAEARKTRSMLSSEDLKLCVLEOYKYLKVCGEDEPLEKYSQYKYLISC
IMGRMNLMMAKESLSYLSOLPMDCFPMPSYSRISTATPMNGESTKSLWYNSL
RIKICATYVNLNRIIDIKYVRGITHGGEPLCDNYNTQVPCSNRMENLNDYI
IPDLPRARCLSLSCYKGRKGAKEEPLAMGINTLPDITDILVSGMAMLVNPFVH
GLEDLNPIDVGTSPNPKETPCLELEDFWSSVVKFPMVSEHNAWSREAGFESY
SHAGLSRLARNDELRENDKOLRAICTRDLSEITDEKDFLMSHRHAYCVTIPILP
KLIVSVKNSRDEVAQMYCLVKDMPPIKPEQAMELLDCNPDPMVRFGEAVCEKYL
DDKLSOYLIDLVYLYKYOYLNDNLVRFELKALTNORIGHFPMHLSKEMHNTVSO
RFGLLSYSCBACGMYLKHNRYEAKELINLIDIKORKEDETKYOKKFEVYEQNR
IPFNSGDLQODMLTIOITRIMENIMQNGDLMLPYGLSIGDCVGLLEVYNSHT
IMJOCKGKLGALQFNSHTLHOMLKONKEITDAIDLETSCACVATFVLIG
DRHNSNIMVDDQGLFHDGFHLDHKKRKFGRKREVPFLVODFLIVISKGOECT
KTRFEFEQECYKAYLAIRQHANLFTNLSMMLGSGMPELQSPDDIAYIRKTLALDK
TEOALVEPKOMNDAHGGWTTKMDIFRTHKOHALN"

BASE COUNT 1134 a 618 c 709 g 963 t
ORIGIN

Query Match 100.0%; Score 19; DB 10; Length 3424;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGCTTCTTAGCATCA 19
|||||
Db 890 AGGCTTCTTAGCATCA 872

RESULT 5
HSU79143/c 3207 bp mRNA PRI 02-JAN-1997
LOCUS HsU79143
DEFINITION Human phospholipase 3'-hydroxykinase p110-alpha subunit mRNA,
complete cds.
ACCESSION U79143
VERSION 94763625
KEYWORDS U79143.1 GI:1763625
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3207)
AUTHORS Stirdivan,S.M., Ahern,J., Conroy,R.R., Barnett,S.F., Ledder,L.M.,
Olliff,A. and Helmbrook,D.C.
TITLE Catalytic Activity of the p110-alpha Subunit of Human
Phospholipase 3'-Hydroxykinase is Required for Signal
Transduction
JOURNAL Bioorg. Med. Chem. (1996) In press
REFERENCE 2 (bases 1 to 3207)
AUTHORS Stirdivan,S.M., Ahern,J., Conroy,R.R., Barnett,S.F., Ledder,L.M.,
Olliff,A. and Helmbrook,D.C.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs,
Sumneytown Pike, West Point, PA 19486, USA

FEATURES
source
1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3207
/function="phospholipid kinase"
/codon_start=1
/product="phospholipase 3'-hydroxykinase p110-alpha
subunit"
/protein_id="AA839753.1"
/db_xref="PIR:P1763626"
/db_xref="GI:1763626"
/translation="MPRPSGELMGHIMPRLIVECLPNGMIVTLECREATIT
IKHLEFEARKYPLHOLLODESSITFVSVOEARREPEFDETRLCRLFOPLKAT
EPVGNREKILNREIGFALGMPCEFDVNDPEVODFRNRLINVCRAVDRLDINSBH
SRMAYVPRPVSSEPELPHKYNKLDGQIIVYIVWVSPNNOKYTKLNNHCVE
QVIAEARKTRSMLSSEDLKLCVLEOYKYLKVCGEDEPLEKYSQYKYLISC
IMGRMNLMMAKESLSYLSOLPMDCFPMPSYSRISTATPMNGESTKSLWYNSL

RIKLACATYVNVNTRIDIKYVRCIYHGEPLCDNVNTQVPCSNPRNWNLYNDYI
 IPDIPRAKCLLSTICVAGKRGAKBEBPCLPMGNINLBDYDITLVSQWALNMPVPH
 GLEDLNPIVATGSPNKEETPCLELEFDMFSSVFKPDMDSVIEELANMSVREGFSY
 SHAGLSNRLADNDELRENDKBLKAISTRDPLSETEBEKDFLMSHRVYCTIPEILP
 KLLSVKNSRDEVAQMYCLVMDPPIKEPQAMELDCNYPDPWARGAVACLEKLYT
 DDKLSOYLIOLOVYAKYBOYLDNLVRLKALKLNORIGHEFFEMHLSKSMHNKTVSO
 RFGILLESYCRACGMVYKHLNROYAMERKILNLDITLKOEKDETOYQOMPELYDOKR
 RPDMDALQGLFSLPNPAHOJGNLBLECRILMSAKRPLMLNWPDIWSELLOPONE
 IIFRNGDDLRODMTLTLQIIRIMENIMWQGLDRLPYGCLSTIGCVGLIEVRSHT
 IMOJOCCKGLGALGOFSNTHLQWLKDNKEIYDAADLETRSCAGYCAVFTILIGIG
 DRHNSNIMVKDGOLEFHDGFLDKKKKREYKREYFVLTODFLIVISKGOECT
 KTREREOEMCKAYALIAIROHANLEINLFMSMLGSGMPELOSFODIAYIKRTALDK
 TEOELEFEMKQMDAHHGWTWKMDWIFHTIKOHALN*

BASE COUNT 1043 a 586 c 670 g 908 t
 ORIGIN

Query Match 100.0%; Score 19; DB 10; Length 3207;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGGCTTCTTAGCCATCA 19
 Db 878 AGGCTTCTTAGCCATCA 860

RESULT 6
 AF001075 3389 bp mRNA VRL 08-JUL-1997
 LOCUS Avian sarcoma virus 16 gag-v-phosphoinositide 3-kinase catalytic
 DEFINITION subunit fusion protein (gag-v-p3k) and retroviral env mRNAs,
 partial cds.
 ACCESSION AF001075
 NID 92245502
 VERSION AF001075.1 GI:2245502
 SOURCE Avian sarcoma virus 16.
 ORGANISM Avian sarcoma virus 16.
 REFERENCE 1 (bases 1 to 3389)
 AUTHORS Tsichlis, P.N., Cantley, L.C., Roberts, T.M., and Vogt, P.K.
 TITLE Transformation of chicken cells by the gene encoding the catalytic
 subunit of pi 3-kinase
 JOURNAL Science 276 (5320), 1848-1850 (1997)
 MEDLINE 97334438
 REFERENCE 2 (bases 1 to 3389)
 AUTHORS Chang, H.W., Aoki, M., and Vogt, P.K.
 TITLE Direct Substitution
 JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
 Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
 BCC-239, La Jolla, CA 92037, USA
 FEATURES
 source 1. 3389
 /organism="Avian sarcoma virus 16"
 /db_xref="taxon:60629"
 <1..3389
 /gene="gag-v-p3k"
 /note="fusion gene of gag and v-p3k"
 1..78
 /gene="gag-v-p3k"
 /note="derived from gag gene"
 <1..3252
 /gene="gag-v-p3k"
 /codon_start=1
 /product="gag-v-phosphoinositide 3-kinase catalytic
 subunit fusion protein"
 /protein_id="AAB6533.1"
 /db_xref="PID:g2245503"
 /db_xref="GI:2245503"
 /translation="QQLIRTPASTLTTPGELIKYVLDRKASHMLPRILVECLPVG
 MIVTEGCRKATLTIRKELKAKKPYLDLQDESSYISVTOAEAREFEDETR
 RLCDRLFQPLKVIKIEPVGNEKILNKIGFAIGMPCICEIDMVKDPEVDQFRNIIIN

CDS
 misc-feature
 gene
 misc-feature
 CDS

VCKEAVLDRDANPHSALTYVCPNVSSEDELPKHTYINKDKOIIYIVWYSPND
 KQRTTLKINHCVPEQVYIARIRKTKSMILSSQDLVLEVOGKILVYCGDETL
 LEKPLQYKIRISCIIMGLNPMNAKSLVYQLDPLDVEYTSYRSTANPYAN
 GEATKALWTLINSALIRILICATVNVNIDIKYVTOIYHGEPLCDNVNTQVPH
 CSNRMENISYDIPDLPPRAALSTICSVKRGAKBEBPCLPMGNINLBDYDITLVSQWALNMPVPH
 LVSGRMLNMAVPHGLEDLNPIVATGSPNKEETPCLELEFDMFSSVFKPDMDSVIEELANMSVREGFSY
 EHWMTISREGEFVSTAGLSNRLADNDELRENDKBLKAISTRDPLSETEBEKDFLMSHRVYCTIPEILP
 WSHRHCYKTPPEILPKLLSVKNSRDEVAQMYCLVMDPPIKEPQAMELDCNYPDPWARGAVACLEKLYT
 MVRAPFAVCEKELYTDKLSOYLIOLOVYAKYBOYLDNLVRLKALKLNORIGHEFFEMHLSKSMHNKTVSO
 FVHLKSMHNKTVSOQFGLLESYCRACGMVYKHLNROYAMERKILNLDITLKOEKDETOYQOMPELYDOKR
 ETQVOKKFLVEQQRDPDMALOGFISPLNPAHOJGNLBLECRILMSAKRPLMLNWPDIWSELLOPONE
 ENPIMSELVEYRSTITTIQKGLKRALDPLNSHTLQWLKDNKEIYDAADLETRSCAGYCAVFTILIGIG
 GDGCLIEVRSHTIMOJOCCKGLGALGOFSNTHLQWLKDNKEIYDAADLETRSCAGYCAVFTILIGIG
 SCAGYCAVFTILIGIDRHSNIMVKDGOLEFHDGFLDKKKKREYKREYFVLTODFLIVISKGOECT
 ODLLIVISKGOECTKTREREOEMCKAYALIAIROHANLEINLFMSMLGSGMPELOSFODIAYIKRTALDK
 FDDIAYIRKTLALDKTEOELEFEMKQMDAHHGWTWKMDWIFHTIKOHALN*

BASE COUNT 1071 a 655 c 751 g 912 t
 ORIGIN

Query Match 100.0%; Score 19; DB 17; Length 3389;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGGCTTCTTAGCCATCA 19
 Db 923 AGGCTTCTTAGCCATCA 905

RESULT 7
 M003279/c 3207 bp mRNA ROD 25-MAY-1994
 LOCUS Mus musculus Balb/c phosphatidylinositol 3-kinase 110 kDa subunit
 DEFINITION mRNA, complete cds.
 ACCESSION U03279
 NID 9414994
 VERSION U03279.1 GI:414994
 SOURCE house mouse.
 ORGANISM Mus musculus.
 REFERENCE 1 (bases 1 to 3207)
 AUTHORS Klippel, A., Escobedo, J.A., Hirano, M., and Williams, L.T.
 TITLE The interaction of small domains between the subunits of
 phosphatidylinositol 3-kinase determines enzyme activity
 JOURNAL Mol. Cell. Biol. 14, 2675-2685 (1994)
 MEDLINE 94167738
 REFERENCE 2 (bases 1 to 3207)
 AUTHORS Klippel-Giese, A.
 TITLE Direct Substitution
 JOURNAL Submitted (05-NOV-1993) Anke Klippel-Giese, CVRI and Dalichl
 Research Center, University of California San Francisco, 505
 Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
 FEATURES
 source 1. 3207
 /organism="Mus musculus"
 /strain="Balb/c"
 /db_xref="taxon:10090"

misc-feature
 misc-feature
 3'UTR
 CDS
 misc-feature
 misc-feature
 CDS

CDS

1. 3207
 /standard_name="p110 protein"
 /codon_start=1
 /product="phosphatidylinositol 3-kinase 110 kDa subunit"
 /protein_id="AA18334.1"
 /db_xref="GI:414995"
 /db_xref="GI:414995"
 /translation="MPPRSSGELMGHLMPPRLVBCLEPNGMIVILECRLATVY
 ITHLEFPAKRYPLHQLQDETSTIFVSVQEAREREFEDETRELCDRLQPELKYI
 EPVGNREKILNREIGVIGMPVCEFDVMDQEDERNILNVCKEAVDRLDPSH
 SRMYVVPVNESPELPKHYKRLDKQLIVLVIVSPNNNDQKYLKINHCEVE
 QVIAETKRTKMSLSEQLCVLEQGGYIKVGCDEYFLEKYSQYKTIKRS
 ILMGRPMIMAKESLSYSLPDSFPMSPRSRTSLTAPYMGSTSRKSIWVNSAL
 RIKLICATYVAVNRDIDKTYRGTGTHGGPCLDQNVNTQVPSNPMMWMTDIT
 IPDLPLRLALCLISVKGKRGAEHCPLWAGNI NFDYDITLVSGKMLNLMPVPR
 GLEDLNPLIGVTSNPKETPCLEFEDWFSVVKFEDYVLEANSVSREAGFSY
 SHTGLSNLARDNELRENDKEQLRCLRDLESETBEKDFLMSHRYCVTIEILP
 KILTSVKNSSDEVAOMCYKDMPIRQAMELDQNPDPVAVRSFAVCLERKLT
 DDKLSOYLIOVYLYKYOYDNLVPLRLKALINORIGHFFPMHLKSEHNTVSQ
 RFGILLESYCAGCMYLNKLNROYAKELNLNIDLEQEKDETQKQKFLVEQK
 QPDMALQGLSLSPNPAHQNLRLSECRIMSSAKRPLMNMENPDIMSELQNM
 IIFNNGDRLQDMTLTIIRIMENIMQNGDLNMLPYGLISGCVGLIEVNSHT
 IMQIOCKGGLKGLQFNSHTLHQLKDKNGEIDYDAIDFTSCAGCVATFLLIG
 DRHNSNIMVKDQGLFDHDEGHFLDHKKKKRKYRERYFLTQDELIVISKGOEY
 KTRFEERQECYKAYLAIRHNLFLTFESMGLSSGPELQSFDDIAYIKITLALDK
 TEQALEFTIKQNDADHGGWTFYMDWLFHTIKQHALN"

BASE COUNT 958 a 693 c 734 g 822 t
 ORIGIN

Query Match 91.6% Score 17.4: DB 12: Length 3207;
 Best Local Similarity 94.7%: Fred. No. 26;
 Matches 18: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCTTCTTACCATCA 19
 |||||
 Db 878 AGGCTTCTTGGCATCA 860

RESULT 8
 HS1068E13/c
 LOCUS HS1068E13 161222 bp DNA HTG 11-JUN-1999
 DEFINITION Homo sapiens chromosome 20 clone 1068E13, WORKING DRAFT SEQUENCE,
 in unordered pieces.
 ACCESSION AL035563
 MID 95051828
 VERSION AL035563.15 GI:5051828
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 161222)
 BATES, K.
 Direct Submission
 Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humgeny@anger.ac.uk Clone requests: clonerequest@anger.ac.uk
 On Jun 12, 1999 this sequence version replaced gi:4902588.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress
 and the release of this data is based on the understanding that the
 sequence may change as work continues. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage etc. Order of segments is not known; 800 n's separate
 segments. Unfinished: dj1068E13 Contig_ID: 00053 acc=AL035563
 Length: 1949 bp Unfinished: dj1068E13 Contig_ID: 00899
 acc=AL035563 Length: 1412 bp Unfinished: dj1068E13 Contig_ID:
 02208 acc=AL035563 Length: 1145 bp Unfinished: dj1068E13
 Contig_ID: 01831 acc=AL035563 Length: 1105 bp Unfinished:
 dj1068E13 Contig_ID: 00957 acc=AL035563 Length: 148176 bp
 Unfinished: dj1068E13 Contig_ID: 02504 acc=AL035563 Length: 1333
 bp Unfinished: dj1068E13 Contig_ID: 02342 acc=AL035563 Length:...

1302 bp.
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 Source location/Qualifiers
 1. 161222
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="1068E13"

BASE COUNT 48587 a 33616 c 32167 g 42049 t 4803 others
 ORIGIN

Query Match 84.2% Score 16: DB 34: Length 161222;
 Best Local Similarity 100.0%: Pred. No. 1.2e+02;
 Matches 16: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCTTCTTACCATC 18
 |||||
 Db 44809 GCTTCTTACCATC 44794

RESULT 9
 AE001605
 LOCUS AE001605 10553 bp DNA BCT 22-FEB-1999
 DEFINITION Chlamydia pneumoniae section 21 of 103 of the complete genome.
 ACCESSION AE001605 AE001363
 ACCESSION 94376455
 MID 94376455
 VERSION AE001605.1 GI:4376455
 KEYWORDS

SOURCE Chlamydia pneumoniae.
 ORGANISM Chlamydia pneumoniae

REFERENCE 1 (bases 1 to 10553)
 Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.

AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
 Grimwood, J., Davis, R.W. and Stephens, R.S.
 Comparative genomes of Chlamydia pneumoniae and C. trachomatis

JOURNAL Unpublished
 TITLE 2 (bases 1 to 10553)
 AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
 Grimwood, J., Davis, R.W. and Stephens, R.S.

JOURNAL Direct Submission
 Submitted (01-DEC-1998) Program in Infectious Diseases, University
 of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA

FEATURES location/Qualifiers
 1. 10553
 /organism="Chlamydia pneumoniae"
 /strain="CWL029"
 /db_xref="taxon:83558"

255. 1427
 /gene="Cpn0186"
 255. 1427
 /gene="Cpn0186"
 /transl_table=1
 /codon_start=1
 /product="similarity to Ctl19 Inca"

/protein_id="AA018339.1"
 /db_xref="PIR:G4376456"
 /db_xref="GI:4376456"

/translation="MSRPVNTNPAPNIPPIPTPTGIPPTKRSSEIERYIVAKYI
 LFAIATAGSALGILGALPGIAGLVIIFVSVVLGLILKDSISGSEERLE
 EVSRTEENORLVITTLTETVEVDLKAQDLLEAFENENGNLTKADEEDV
 SKLSEQLERINQIOLANAGDOESLTKLISGMDSKVVOINTSIOALAKVLE
 QEWQKQYTRKNAQEOIQANQATILGHNSSTALQSYENLIVQDALPFAVEELLE
 SENLSDQCSALROETELQHEHSIDQRIQAMLAQONLAEVTALEKMQDAQAE
 SEFIACVADRTRFGRRETPPTTPEVGEDESGEDGTPPVSPSSPVDRATGQG"

1604. 2404
 /gene="Cpn0187"
 1604. 2404
 /gene="Cpn0187"

/codon_start=1

gene
 CDS

/transl_table-11
/product- "predicted methylase"
/protein_id- "AAD18340.1"
/db_xref- "PID:94376457"
/db_xref- "GI:4376457"
/translation- "MHSKPLSRKRRKNSHKEETSMDCIASYNNKIYODKGHYHRETI
LPOLLSTLIGSKSVLDICGCGFLERAPKRCRILGIDISSRLALAKKMSVNSH
ORVADLSKRIEVEPTLFSHAVALSLQWMEPEGHINTATLLELGGOFFLVNHP
CFRIIPRASSWHYDENRKALSRHIDRYLSPKIPIDMAPGGQDPSPTLSFHPPLSYWFK
ELSSHGLVLSGLEWTSKSTGTGRAKAENICRKEPPLFLMISCIKIK"
2467..3759
/gene- "Cp00188"
2467..3759
/gene- "Cp00188"
/codon_start-1
/transl_table-11
/product- "Ct132 hypothetical protein"
/protein_id- "AAD18341.1"
/db_xref- "PID:94376458"
/db_xref- "GI:4376458"
/translation- "MFRKLFPSKKKTGQRQRNLNNGLLQAIQISIKVLLHNASKEA
CVLSYGLVCAPLIVFELRSQHEFTNLNKKEMLIKEPDYKRPVAVAEVNAHTE
SNIGLVGSEFVEFWAGIMLSLSDGLNKIRTSWPISTLRKLSYVITLVSAMI
FIIVGSMYITQIMPIQYAKFLPSLSHSMATLFSRFPYLLLYLALFCCYALFRV
AIQKTSALISTLIGSVWIVFQKAFPSLOVSPNSPTYGALVALPSPILLIYITMI
YLFGALFTFIONRGTFTPLGDKIIPSCILQITSTYITLATTROPNEGLSLTQF
IAKQSVPIGEVSCUDVLEKEGFLPPYNGYQPVNFSELTIKDIAIDLHREIFRK
FNPDLGITFIEVNFQNIENFQASKNKENTLSEIARIRIK"
3756..7184
/gene- "Cp00189"
3756..7184
/gene- "Cp00189"
/codon_start-1
/transl_table-11
/product- "Ct131 homolog (Possible Transmembrane Protein)"
/protein_id- "AAD18342.1"
/db_xref- "PID:94376459"
/db_xref- "GI:4376459"
/translation- "MKRRWKLKLGICLASSYVGLFLLPQLSTESKRYLVESLIH
KESGLSCSAEELKISWFGQTAARKILTEGAKDEVSAEKFELDGLRLILYKRRKG
ITLWSGLKINEPASIDHPSVSHDPGSLTYLNDCKIISEHGFITMKTVSGSLSVS
GYLSEKSEKFNKCVSEDOQGNIFIESVLSIPAQFSAQFVPAEFKPIADPF
MHLISYEDILINSAETHRNDGKISMTASGEGNOIOMKLOGHIRKSTFYVSGSSP
TELPELASALCNOIIPSTPISKOIHATYSARAKIPDITKRIKHEITSQALPEVA
IHRKPNALAQDLNKLGIKTEKESDIRSSSTVLAGSPSHLNGLISIDNKHILK
FRQAOQLPHTYLRALFPOPEVINVPLDAVYSLNLEGYKNAHLEADAILDMPKFLK
SCMSGTWKNFLEKGGQTYHFNKMOEILSPHSYAEARFGSKAQTIDNLFPPKFG
KITARENELLIHAKFGSPNEPIKETTSILIHGQFCSLPLSVSNHLAEFLHLKTFE
FHTDGGKFTYKGNLQALINPOYDPLNNTIRILIDLLSLDESSTSPSKDLKIOGG
ETISLPLDSTTKYKQVRUSPEFGSSGOLNPPVANNPDQNKTLILSNFKSALIGE
LKIYDPSKMLSSGTOGITQWVEYSPERISAFKNAASCSPCLLHRTANRLDLSKLC
PEETKGLSCLTTLAAGLEGSLAETAPLIFYDWSKETFIINDEKSLRANNLDAKLEY
DLKQSLAPRODSKTLAEFSLBEGVDHLESPSEKREPKQAMLIHPSSEIAGIIPASP
GKAQSIISLAGPRINIVSINKNAFEGEPDQIMWSNLOQAILIHINSEKSLIRENLT
AHLSTINEDKAFIAEPLNLAGAYSOYPTVLEIDKONFYPIRYSDEEPFISQAT
LHGKTSIANTGTMVALROFLDITDQKQVESFETPIFSVQGSITCKRLDALIR
IRALMGKTDIANDRLPMTUGIDPEVITKTFHTSILKTNFELIKIRGSIISSEVDWS
SAVARIALKTSYSLGNPSSLADKLSSLDGSTPPPTVHPFMEKSNFDSIENK"
7275..8633
/gene- "Cp00190"
7275..8633
/gene- "Cp00190"
/codon_start-1
/transl_table-11
/product- "hypothetical protein"
/protein_id- "AAD18343.1"
/db_xref- "PID:94376460"
/db_xref- "GI:4376460"

SPPLPGLTSGISGTFWKSASKNSNGSPFSLARKKETESDIDCOTITSTYSGNAG
TTWLSLKVLSVPSIFQIERPEVQSLVSYEDWLPIDNIEMFSQPRITPLALQOTM
LGOXYDILDELAAHQINOTLMTISPNCRSRSLQKOTNOFENSPVDIFYVAHAHSCHWS
GF"
/complement(8667..9347)
/gene- "glnP"
/note- "Cp00191"
/complement(8667..9347)
/gene- "glnP"
/codon_start-1
/transl_table-11
/product- "ABC Amino Acid Transporter ATPase"
/protein_id- "AAD18344.1"
/db_xref- "PID:94376461"
/db_xref- "GI:4376461"
/translation- "MTIRRNLAYSVNKKILIDGVFSLERHITLTFVKGSGSGKMTI
LKAALGVPTGDIWIEGEAPALVQGELESHMTVGNCTHPQIHIGKRSERARE
KAPELIHLDIEVAKNPDOLSGGOKQVAVRSLQMDKHTLLFDEPTSAIDPEPATA
SRHLEETIRDELIVGLTTHDMQVYHSCLDRIYILIDGTAVGYDKRDELDSGHL
SKYHSAQ"
/complement(9337..9990)
/gene- "glnP"
/note- "Cp00192"
/complement(9337..9990)
/gene- "glnP"
/codon_start-1
/transl_table-11
/product- "ABC Amino Acid Transporter Permease"
/protein_id- "AAD18345.1"
/db_xref- "PID:94376462"
/db_xref- "GI:4376462"
/translation- "MDHWATARLLRGGCTYLCVSGIGIIGSLIGLIGVTSLYF
PSKLTRLANSIVTVIRGTPLEFLOLITLYFGLPEVLPYEPPLVAGIILASVLSAYL
AENIRGINSLSIGQESAMVLYGKYQIIFYVYIYQVFNKILPSLIEPVSLSKISS
ILMVGVPELTIVTRDIVERSREINPMENYLICAGLFLMTSPSCISRLSEKRSYDN"
/complement(10039..10482)
/gene- "argR"
/note- "Cp00193"
/complement(10039..10482)
/gene- "argR"
/codon_start-1
/transl_table-11
/product- "Arginine Repressor"
/protein_id- "AAD18346.1"
/db_xref- "PID:94376463"
/db_xref- "GI:4376463"
/translation- "MKKRVTTIDEALKEILLREGAATQEBLCAKLIAQFATQSSVS
WLRIOAVVAEGRCARSLPSESTKTRHLVLSIRHNASLIIVRTVPGASMTAAL
LDQKIDELIGTAGDDPIFVPRIDEGSLPLMWSIANLLQVFLD"

BASE COUNT 3353 a 2292 c 1802 g 3106 t

ORIGIN

Query Match 83.2% Score 15.8; DB 2; Length 1053;
Best local similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTTAGCCATCA 19
|||||
Db 3188 AGGCATCTTTAGCCTTCA 3206

RESULT 10
L1007640/c L1007640 1138 bp DNA BCT 05-JAN-1995
LOCUS L1007640 1138 bp DNA BCT 05-JAN-1995
DEFINITION L1007640 Lactococcus lactis triosephosphate isomerase (tpl) gene, complete cds.
ACCESSION U07640
NID 9538127
KEYWORDS U07640.1 GI:538127
SOURCE Lactococcus lactis.
ORGANISM Lactococcus lactis

Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

Lactococcus.

REFERENCE 1 (bases 1 to 1138)

AUTHORS Cancelli,M.R., Davidson,B.E., Hallier,A.J., Nguyen,N.Y. and Thompson,J.

TITLE The Lactococcus lactis triosephosphate isomerase gene, tpi, is

monocistronic

JOURNAL Microbiology 141 (Pt 1), 229-238 (1995)

MEDLINE 95202084

2 (bases 1 to 1138)

AUTHORS Davidson,B.E.

TITLE Direct Submission

JOURNAL Submitted (10-MAR-1994) Barrie E. Davidson, University of

Melbourne, Biochemistry, Parkville, Victoria, 3052, Australia

On Sep 15, 1994 this sequence version replaced gi:537285.

COMMENT Location/Qualifiers

FEATURES

source

1..1138

/organism="Lactococcus lactis"

/sub_species="lactis"

/isolate="LM0230"

/db_xref="taxon:1358"

/clone_id="lambda GEM11"

/map="1.818 to 1.978 Mbp"

72..77

102..107

115..988

/evidence="experimental"

RBS 165..170

gene 178..936

/gene="tpi"

178..936

/gene="tpi"

/ec_number="5.3.1.1"

/codon_start=1

/transl_table=1

/evidence="experimental"

/product="triosephosphate isomerase"

/protein_id="AAC3268.1"

/db_xref="PID:9537286"

/db_xref="GI:537286"

/translation="MSRRPITAGMMKMTLEAQAFAEAVKNNLPSSDNVESYIGAP

ALFLAPAYLRGSEELKLAENSYFENAGFTGNSPAIVDGLIYIIIGHSRRREY

PHETDEDINKKAKAIFPAAGATPICCGSETLTFRGKTAEMVSCQTEGLAGLTAEOY

SNLVIAETPAIETGKATNEIADETGQVRSYVENLTKGEVSEAVRIQYGGSVKPE

TIETMAKENIDGALVGASLEADSFLLLEMYR"

958..985

/note="rho-independent terminator"

/function="putative transcription terminator"

/rpt_type="inverted"

/rpt_unit="958..970

958..988

terminator

BASE COUNT 379 a 192 c 235 g 332 t

ORIGIN

Query Match 83.2%; Score 15.8; DB 2; Length 1138;

Best Local Similarity 89.5%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AGGCTTCTTAGCCATCA 19

DB 863 ATGTTCTTAGCCATCA 845

RESULT 11

LOCUS D86747 3060 bp mRNA VRT 07-FEB-1999

DEFINITION Chicken nei mRNA for Mr 93 k protein (with EGF-like repeats),

complete cds.

ACCESSION D86747 DA5365

NID 91483183

VERSION D86747.1 GI:1483183

KEYWORDS Mr 93 k protein; nei.

SOURCE

ORGANISM

Gallus gallus

REFERENCE

AUTHORS

1 (bases 1 to 3060)

TITLE

JOURNAL

Direct Submission

Submitted (26-JAN-1995) to the DDBJ/EMBL/GenBank databases. Sachiko

Matsunashi, Saga Medical School, Department of Biochemistry, 5-1-1

Nadeshima, Saga, Saga 849, Japan

(E-mail:matsunash@post.saga-med.ac.jp), Tel:(0952-31-6511(ex.2265),

Fax:(0952-33-2511)

2 (sites)

Matsunashi,S.

REFERENCE

AUTHORS

3 (bases 1 to 3030)

TITLE

JOURNAL

Dev. Dyn. 203 (2), 212-222 (1995)

COMMENT On Aug 8, 1996 this sequence version replaced gi:642136.

LOCATION/Qualifiers

FEATURES

source

1..3060

/organism="Gallus gallus"

/db_xref="taxon:9031"

/dev_stage="9 day old embryo"

118..2568

/gene="nei"

118..2568

/gene="nei"

/note="with EGF-like repeats"

/codon_start=1

/product="Mr 93 k protein"

/protein_id="BAA13167.1"

/db_xref="PID:d1013856"

/db_xref="GI:1483184"

/translation="MESGGCGIGTCLICLCIGPVYVPGVDSIQTDVLSGLTSPYAG

VROYPLGHSKAPLFPDTSRSVASPELAEIIFOKLRNKYEFTITLYLKOAHNSGV

IFSTHLDHRYLLESSGHRNEIRLHRTGSRSHTEVPYLLADDDWHRLSLAISAS

HLIHLVDCNKIYERVENKPEMDLPGTFYMGQNNMAGYFKGMQVOLIYVQGLI

SCQEDLNRCPTCDNFGLVOKIMELDDIILAKTSAKSOAQRNKKIDOCVCEYCN

KGMVREPESEWTDCKNCTGNGVQCEALICSLDQPNALSIVYGGKKCKEQSYC

IFESRTYFEGGRERYSSGDCVLFPEKNDHMQRIIPDSCATLNCPSQOIRLSHSC

KICGHDPCTEGHNCMEHSVCRNLDRAVSCRDGFRLREDNAVCEVDECAEGOH

CRENTMCVNTPGSPMCICIKTGYIRIDYSCDEHDECVTNHNCDEMLCFETVGHNC

VCKLGYNGVVCACAFKPCDGRNGAGIASVACAPGFTGSPSCETIDIDECSEFYVC

DSRANGLNLPQWYCECRDGYHDNGFSPSGSEDEDIDECATGHRSCANDVYCNLNG

GYDCRCPHKNCTGDCIHEDKIRHNGOIVWENRDCRGVSCGSGYVMCRRMYCCEMP

TYDLCPCPECDPRSLSSQCLHRSGLSINSQSYQNCQCCCLQGEVDQNLPEPEVD

CERSVLPENBCPCRCVTDPCQADTIRNDITKCTLDETNVAFETSSWIKHGTCTLCO

CKNGHVCSSVDPQCLQEL"

BASE COUNT 830 a 672 c 779 g 779 t

ORIGIN

Query Match 83.2%; Score 15.8; DB 4; Length 3060;

Best Local Similarity 89.5%; Pred. No. 1.8e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AGGCTTCTTAGCCATCA 19

DB 571 AGGCTTCTTAGCCATCA 589

RESULT 12

LOCUS GGFMYLCK/C 7631 bp mRNA VRT 17-FEB-1997

DEFINITION G.gallus mRNA for fibroblast myosin light chain kinase.

/citation=[5]
 /replace="c"
 5037
 /citation=[5]
 /replace="c"
 5248, .5721
 /codon_start=1
 /product="KRP/Telokin"
 /protein_id="CAA37058.1"
 /db_xref="PID:g982995"
 /db_xref="GI:992995"
 /db_xref="SWISS-PROT:P11799"
 /translation="MANISGMSGRKASGSSPTSPINDKYNEDAPLEAYAEKPHK
 PFTKTLIDHEVYEGSARNDCKIEGTPDEVMATKDDPKESRHFQIDYDEGNC
 LTISEVCGDDAVYTCVANSLSGATCTALLETVMKEGEGEGEDEDEEE"
 BASE COUNT 2435 a 1569 c 1793 g 1834 t
 ORIGIN

Query Match 83.2%; Score 15.8; DB 4; Length 7631;
 Best Local Similarity 89.5%; Pred. No. 1.8e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTACCAATCA 19
 ||||||||| |||||
 Db 1474 AGGCTTCTTACCAATCA 1456

RESULT 13
 GMLCK08 253 bp DNA VRT 08-AUG-1998
 LOCUS Gallus gallus smooth muscle/non-muscle myosin light chain kinase
 DEFINITION gene, exon 8.
 ACCESSION AF045262
 NID AF045262
 VERSION AF045262.1 GI:3403177
 KEYWORDS
 SEGMENT
 SOURCE
 ORGANISM
 * 8 of 31
 Gallus gallus
 chicken.
 Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria;
 Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 1 (bases 1 to 253)
 Wattersen,D.M., Collinge,M., Lukas,T.J., Van Eldik,L.J.,
 Birukov,K.G., Stepanova,O.V. and Shirinsky,V.P.
 Multiple gene products are produced from a novel protein kinase
 transcription region
 FEBS Lett. 373 (3), 217-220 (1995)
 2 (bases 1 to 253)
 2 (bases 1 to 253)
 Birukov,K.G., Schavocky,J.P., Shirinsky,V.P., Chibalina,M.V., Van
 Eldik,L.J. and Wattersen,D.M.
 Organization of the genetic locus for chicken myosin light chain
 kinase is complex: multiple proteins are encoded and exhibit
 differential expression and localization
 J. Cell. Biochem. 70 (3), 402-413 (1998)
 3 (bases 1 to 253)
 Birukov,K.G., Schavocky,J.P., Shirinsky,V.P., Van Eldik,L.J. and
 Wattersen,D.M.
 Direct Submission
 Submitted (30-JUN-1998) Molecular Pharmacology and Biological
 Chemistry, Northwestern University Medical School, 303 E. Chicago
 Ave. Ward-8-196, Chicago, IL 60611, USA
 Location/Qualifiers
 1..253
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /tissue_type="liver"
 /dev_stage="adult"
 27..230
 /number=8
 BASE COUNT 80 a 46 c 50 g 67 t

ORIGIN
 Query Match 83.2%; Score 15.8; DB 4; Length 253;
 Best Local Similarity 89.5%; Pred. No. 2.1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTACCAATCA 19
 ||||||||| |||||
 Db 146 AGGCTTCTTACCAATCA 128

RESULT 14
 AC007258/c 144422 bp DNA PLN 16-JUN-1999
 LOCUS Arabidopsis thaliana chromosome I BAC F23H11 genomic sequence,
 complete sequence.
 DEFINITION
 ACCESSION AC007258
 NID AC007258
 VERSION AC007258.3 GI:4733952
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 thale cress.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euembryophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 1 (bases 1 to 144422)
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
 Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
 Gonzalez,A., Kremetska,I., Kim,C., Lenz,C., Li,J., Liu,S.,
 Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.S.,
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.M.
 Unpublished
 2 (bases 1 to 144422)
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
 Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
 Gonzalez,A., Kremetska,I., Kim,C., Lenz,C., Li,J., Liu,S.,
 Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.,
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.M.
 Direct Submission
 Submitted (04-MAY-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 3 (bases 1 to 144422)
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
 Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
 Gonzalez,A., Kremetska,I., Kim,C., Lenz,C., Li,J., Liu,S.,
 Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.,
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.M.
 Direct Submission
 Submitted (16-JUN-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 4 (bases 1 to 144422)
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
 Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
 Gonzalez,A., Kremetska,I., Kim,C., Lenz,C., Li,J., Liu,S.,
 Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.,
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.M.
 Direct Submission
 Submitted (16-JUN-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA

COMMENT
 On May 4, 1999 this sequence version replaced gi:4678191.
 Bases 138041-144421 'IGF' Clone F23H11 overlap with bases 1-6382 of
 'TAMU Clone T2K10', Accession AC005396.
 e-mail for correspondence: arabssequence.stanford.edu
 Genes with similarity to proteins in the databases are described
 as 'putative', '-like' or 'similar to'. Genes that have EST
 similarity but no significant protein similarity are described as
 'unknown proteins'. Genes that are annotated based only on gene
 predict ion software are described as 'hypothetical proteins'.

The software programs used to predict genes include: Graftl (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge, <http://genome.stanford.edu/~chris/GENSCANW.html>), Fera (V. Solovjev & A. Salamov, Sang er Centre, <http://genome.sanger.ac.uk/>), Neplantgene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/Neplantgene.html>) and EMOtiff (Nevill Manning, C.G. Wu, T.D. & Brutlag, D.L., <http://motif.stanford.edu/projects.html>).

FEATURES

```

source
1..144422
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/clone="F23H11"
/chromosome="1"
2303..3607
/gene="F23H11.1"
join(2303..2629,3230..3607)
/note="Similar to glutathione transferase"
/codon_start=1
/protein_id="AAD39312.1"
/db_xref="PID:95080803"
/db_xref="GI:5080803"
/translation="MGKEEYKLGWVSPYAIRKIALRLKSDYDYEENLRGSKS
ELLKSNPEYHKVPYLHNKPIVESLIVETIDETNNSSAPSLIPSPYDALRWM
SDVDNKEFPALRMAIIRKSDAKAKAEVEEGLEDAFVLSISKRPFGGAIG
FMIDICFSGFVLLKAREKFKAEKLEDESKTPSLCKMADRFLSDETKNAVEIEKVAE
FLOLEVEAQAQASASRS"
4467..5517
/gene="F23H11.2"
join(4467..5051,5122..5416,5495..5517)
/note="Hypothetical protein"
/gene="F23H11.2"
/codon_start=1
/protein_id="AAD39313.1"
/db_xref="PID:95080804"
/db_xref="GI:5080804"
/translation="MEIFOKAKAVALLRSHDKYLVADEESVTQERNSGAAMWTY
EIIPIGSTWILRLKSAKYGLYASNKPILGATGKVLQTKPSRLDSSLAAMPISDHAL
VKIKTRYGNFLKNGSLPMPKNSVTHDIPHSATQCEVLMHIDVLEILPANAQSDSH
OHLQLOQPSPLHSDSLDPTPGSPSRSDREFRQSTDSLAVGSPKSGEIVTYHVA
DDDDVEDDSVEVSFTKNGVEELTKLKEESNVEVYICTRSPNGKLFPLRLQL
PPNADMTVVLVPPSSKIAEKFK"
complement(5712..8207)
/gene="F23H11.3"
complement(join(5712..7462,7970..8039,8181..8207))
/gene="F23H11.3"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39314.1"
/db_xref="PID:95080805"
/db_xref="GI:5080805"
/translation="MEDTKLGDOKSSNVSLGSSCYOLDRYHATVLOTGSDMSOLKOL
HAFILRTTYPEEPATPLPYGKTIQDSSSPDVNAYAFRDEDSIENHSSPMWTLIRACA
HDSRKEAEKMLYRKMLERGSSSPDKRTFFPYLKACATIFPSSEKQVHCQIVAHGEG
GDVYVNNGLILHYSGCGCLDIARVYEDMPERSLVSMSMDALVAFEGEYSALQLEFR
EMORSFEPDGYTQSGCAGLSLGTNAHAEFLKRCVDVADVALVNSLIEMY
CKGSLRMAEQVFGOMKRDLSANMAILGFATGRABEANFEDRMVKKRENPNS
VTFGILACIHNHGFVNGKROYFDPMVADYCIPEPLIEHYGCIYDLIAAGYITTAIDM
VMSMPKPDVAVIWSLIDACCCKGASVLSSEIARNTIGTEEDNESSNGNSGAYLL
SRVYASASRNDVGLIRKLMSHGHGIRKPGGCSLEINISHEFTFGDTSHPOTQIIO
QLKVIDDLRLRIGILPDRSQAPLVADATNGSKEYSLRLHSRLAIARGLINLPOTPI
RIFNLRLCNCDCHEVTKLISKVFYVEIIRDRVRFHHRKDGSCSLDWM"
16582..18054
/gene="F23H11.4"
join(16582..17135,17613..18054)
/note="Very similar to hsp40, heat shock proteins."
/codon_start=1
/product="Putative heat shock protein"
/protein_id="AAD39315.1"

```

```

/db_xref="PID:95080806"
/db_xref="GI:5080806"
/translation="MGVYVYVLANVNPSPATEDDLKRSYRLAMKHPDKNPTSJKOBA
EAKRKQISEADVLSDPNKRQIYQDGEDGLATEAASQOHNYSGGNNYNAAGR
YTPEDADIEFAEFEGASEKVDGCVGGGGRKSAEASQNRKPKVARKPALESLKA
CITLBEIKGGRKKKISRVPDGGKSKPEVLEKIDITGKPKGCTITGPERKNOBP
GVTADLITFYDENKPHYRKDNDLYDKKVSILELMTGITSLSLTLDRNLTIPVL
DIVPGEIVTIPSGMPTSKSGSRGDLRINFELCPFSRLTSEKTDLRVLGSS"
complement(18603..19236)
/gene="F23H11.5"
complement(join(18603..18734,18816..18938,19102..19236))
/gene="F23H11.5"
/note="Similar to thioredoxins. Contains thioredoxin
family motif."
/codon_start=1
/product="Putative thioredoxin"
/protein_id="AAD39316.1"
/db_xref="PID:95080807"
/db_xref="GI:5080807"
/translation="MGSNVSSVDHVSMEITTSNGVYEIESRQKSLSDSKGSK
LVVIDFAVWCGPKAMEPRVREIATSYSAVFAVDVDRIMDAVGYRAITTPAFVF
VKRGEEDRVYGAKPDELVKIKEDHRV"
34973..38156
/gene="F23H11.6"
join(34973..35434,36790..36917,37322..38156)
/gene="F23H11.6"
/note="Similar to nitrate and oligopeptide transporters"
/codon_start=1
/protein_id="AAD39317.1"
/db_xref="PID:95080808"
/db_xref="GI:5080808"
/translation="MYCHYVSOGPIILSVANHLPOLRPKPKPIIDTCEARGFKAM
IFPVALYLVALGSCVPMNIAHADGPSHSPKQSRKLSYEAFAAFASMGELIL
TLVLVQVTHSGMDIGFVSAAMTMGILISVSGMYRNKRPRSITPLAIVYAI
LKRLIASPSDPRMLHGDHVAANDVPSSTLPHPRFPLDKACIKIDTITKESPMIL
CIVQVQVOKTILSLVPIPASTIVENTILMOIGFVQOQSSMTRLSNSFPAPAL
QATPYMLIEFLVPLVDSFIVPEAKLGNHSGIPLRIGIGLIFSNVSAMLEK
KRDSYLDGRILSLITPITPOPLRIGISEPTANGILIEFYKQSAKMESEFALITIC
SYSGFYSSVLVSANRITSTSDVSGMGLENDLNKRDLEFYLAVLSLNFISY
LFWSRNMIKSRNNINVDENT"
46258..50037
/gene="F23H11.7"
/note="ARE1"
join(46258..46294,46500..46615,46871..46966,47073..47129,
47228..47380,47466..47550,47704..47794,47873..48037,
48263..48382,48473..48548,48668..48839,48916..49385,
49532..49722,49869..50037)
/gene="F23H11.7"
/note="Rare gc intron splice site at 47794 is inferred
from alignment with cdna sequence."
/codon_start=1
/product="auxin response factor 1"
/protein_id="AAD39318.1"
/db_xref="PID:95080809"
/db_xref="GI:5080809"
/translation="MAASNSSGKPGGVLDALCRELMAHACAGVLTLPDEGVYF
PEGMEDLESMDHGLEQMPSENLPSKILCKVINIRKREPTEDEYTAITLPELD
QSEPTDAVQVEKCTIVASFCKTLTASDTSGHGSVSKRLVAGDCLPILDSQOP
MOELVAVDLNDSVEHEFHFGPRLRLTIGASVFSKRLVAGDAFIRLGNELT
RVGVARLRROOTNIPSSVISHSMLIVLATAAHATITGIFESVYPRISSEFIVS
VNRLEKTKTQLSGMPFKRPGEEAEPEKFGITVGOENKSSVWHDSEMSITVO
WDESSVFPREPRSPLELELVANSPTSSQOPQPKRRRPRPGLSPATPGSPVTP
DGWKSSTADTPSSVPLESPPAKATFCHGKNKSTGVSIGAFMPTNDSAESFASAF
NNBSETRKQNGNVCRLFGELEVENVWDCFSFAVSVAADQAPVPSVNEPSSQGS
EPLINIOSDILPSSGSDPEKSLSPQSSQORJSCCTKYVMOCSAVGRALDRLRSECY
EDLEFKLEMEIDKDELSTKMKOVYVTTDEDDMAVVGDDPNNEFGVAKRILFYTP
EYVKTLSPKRLAVNAMOIKADAENGNENGSSASGR"
complement(50415..55954)

```

Query Match 83.2%; Score 15.8; DB 8; Length 144422;
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 15
SPAC24H6/c

LOCUS SPAC24H6 36833 bp DNA PLN 22-APR-1999

DEFINITION S.pombe chromosome I cosmid c24H6.

ACCESSION 254142

MID 9984697

VERSION 254142.1 GI:984697

KEYWORDS 40S ribosomal protein S9; 5S rRNA; cdc25; cullin 3; DAHP synthetase family; guanine nucleotide exchange factor; hexokinase; Homol D box; Homol E box; hsk1; M-phase inducer phosphatase; major facilitator family; phospho-2-dehydro-3-deoxyheptonate aldolase; Rhodanese; rps9; ubiquitin activating enzyme; ubiquitin--protein ligase.

SOURCE fission yeast.

ORGANISM Schizosaccharomyces pombe

REFERENCE Eukaryota; Fungi; Ascomycota; Archiacomyces;

AUTHORS Schizosaccharomycetales; Schizosaccharomycetaceae;

REFERENCE Schizosaccharomycetes.

JOURNAL 1 (bases 1 to 36833)

AUTHORS Skelton, J. and Churcher, C.M.

REFERENCE 2 (bases 1 to 36833)

AUTHORS Barrell, B.G., Rajandream, M.A., Walsby, S.V. and Wood, V.

TITLE Direct Submission

JOURNAL Submitted (10-SEP-1995) Schizosaccharomyces pombe chromosome I

COMMENT sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ E-mail: barrell@sanger.ac.uk

Notes:
Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_pombe/)
Protein coding regions (CDS) have been predicted with the help of computer analysis using the Genetinder program in Pombase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.
CDS are numbered using the following system eg SPAC5H10.01c. SP (S. pombe), A (Chromosome 1), c5H10 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid c24H6 is overlapped at the 5' end by cosmid c23E2.

FEATURES

Source

Location/Qualifiers

1. .36833

/organism="Schizosaccharomyces pombe"

/strain="972h-"

/db_xref="taxon:4896"

/chromosome="I"

/clone="cosmid c24H6"

/map="1L"

1

/note="nominal overlap with c23E2 EM:268887"

gene

complement(30..800)

/gene="SPAC24H6.01c"

complement(join(30..478,554..800))

/gene="SPAC24H6.01c"

/note="SPAC24H6.01c, len:231aa; similarity: to YG1084C, YG14 YEAST, HYPOTHETICAL PROTEIN, (560 aa)opt:300, E():1.6e-15, (33.3% identity in 103 aa overlap)"

/codon_start=1

/product="hypothetical major facilitator family protein"

misc.feature

/note="nominal overlap with c23E2 EM:268887"

gene

complement(30..800)

/gene="SPAC24H6.01c"

complement(join(30..478,554..800))

/gene="SPAC24H6.01c"

/note="SPAC24H6.01c, len:231aa; similarity: to YG1084C, YG14 YEAST, HYPOTHETICAL PROTEIN, (560 aa)opt:300, E():1.6e-15, (33.3% identity in 103 aa overlap)"

/codon_start=1

/product="hypothetical major facilitator family protein"

/protein_id="CAA90845.1"

/db_xref="PID:984698"

/db_xref="GI:984698"

/db_xref="SWISS-PROT:Q09758"

/translation="MLRFEDVLESTKTERPNKSSRLSSGSSHPSSRLTV
RSAPVKSAGSIEFLFYSVLIITLACFKIHVSSSPKPKYKNEKXLLKQWLEF
QKYSADPQSAFRENKPIILLVIYINFMRLKLVFTNTNDELAIKNNYLCFSL
LFALVGTGVITVITALLINYLISLKSINPLTLTLDISVVFKEFYAFKFS
SLHPGLGFLVRYI"

complement(479..495)

/gene="SPAC24H6.01c"

/note="splice branch and acceptor sequence,
actactagagcgcag"

complement(348..553)

/gene="SPAC24H6.01c"

/note="splice donor sequence, gtaagt"

complement(1466..2037)

/gene="SPAC24H6.02c"

complement(join(1466..1742,1787..2037))

/gene="SPAC24H6.02c"

/note="SPAC24H6.02c, len:175aa"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAA90846.1"

/db_xref="PID:984699"

/db_xref="GI:984699"

/db_xref="SWISS-PROT:Q09759"

/translation="MMSFARYTKFCRLNQLKNSKRVSSILFCSPVNRISNQ
CKRPTOSDARDIASGVPSKSDADQLQRPITYNSFCVCTNRSNHFSGKQAHNCT
VIVOCPRKRRHIMADLTKFSEBRTIEDILAKGTEFKGIVNGVNERKPPQ
FKIRPAKSSSSNSK"

complement(1743..1759)

/gene="SPAC24H6.02c"

/note="splice branch and acceptor sequence,
tactaccataatgaag"

complement(1781..1786)

/gene="SPAC24H6.02c"

/note="splice donor sequence, gtaagt"

2201..4856

/gene="SPAC24H6.03"

join(2201..2242,2299..3683,3723..3832,3885..3949,
3994..4167,4219..4628,4665..4856)

/gene="SPAC24H6.03"

/note="SPAC24H6.03, len:785aa; similarity: to G3139079,
cullin 3, (768aa), fasta scores, opt:602, E():0, (37.3%
identity in 812 aa overlap)"

/codon_start=1

/product="cullin 3 homolog"

/protein_id="CAA90847.1"

/db_xref="PID:e1312068"

/db_xref="GI:9336937"

/db_xref="SWISS-PROT:Q09760"

/translation="MORSAKLIRAPRFSANOVDPATHEVYLORAGIDFQKSTSL
SFEELYNATILVLHKGKLYNHVQVVISRLKEEVPATVINYDLSLGNALDLIR
KNDYSYSWRSRSLAAHRLPSLSDNSKMSVLYKLYDKVSSAKVPNA
ENGCIYRREVLINSFEIGECVETILLYLEKGMTINRPLINDLMDLNSPNS
KKEIYDVLPAKPLSYTRNEVELESSTVGCVVYLKAEKRFEEKERSNYV
TKIASPLSVDELLSKHDLLENSGTGFSDSNEFGIOLVYESPSELYGK
SLKRYLAKYVAHBEKLTNETSOLAEKMAVGRSSNATATLVQVVLMDRLNTI
ISTTMDRSILNSDAFVTFVQGYRADEYISLFTDDNKRKARAKGISTEATQ
NSVTLPFRISERKLVFEKYRTHLAKRLNRSISDSDELKMSLKQDAVNTQKLE
GMFNDMLSOELLQEYKNSALOSAKPALDNLNSIILASTFMPIDLSPKTKCNPKYL
LAQIDQPTDYLSLKHGKRLKLPMSGSAVRYVFKDKYDLNVSTIASYLLLFQVL
KENOCILFEELIEETKNTIEVGDITNLOSACAKYKILKDKPGREVAAGKRYENEF
VSNARIKISTVATQTRVEDSERKRTIEKYEDESKRHADACIVRYAKDRVCENQLM
AEVTROLNPRFHSPPMIKRIETLIERETLORADNGRIETELA"

2243..2248

/gene="SPAC24H6.03"

/note="splice donor sequence, gtaagt"

2286..2298

/gene="SPAC24H6.03"

/note="splice branch and acceptor sequence, tactaccattag"

Job time: 4586 sec

misc_feature	3833..3838 /gene="SPAC24H6.03" /note="splice donor sequence, gtaagt"
misc_feature	3872..3884 /gene="SPAC24H6.03" /note="splice branch and acceptor sequence, tattaatgcttag"
misc_feature	3950..3955 /gene="SPAC24H6.03" /note="splice donor sequence, gttagt"
misc_feature	3981..3993 /gene="SPAC24H6.03" /note="splice branch and acceptor sequence, tactacaagctag"
misc_feature	4168..4173 /gene="SPAC24H6.03" /note="splice donor sequence, gttagt"
misc_feature	4200..4218 /gene="SPAC24H6.03" /note="splice branch and acceptor sequence, cactacaacccattag"
misc_feature	4629..4634 /gene="SPAC24H6.03" /note="splice donor sequence, gtagtg"
misc_feature	4657..4684 /gene="SPAC24H6.03" /note="splice branch and acceptor sequence, tactacaagctaaataag"
misc_feature	5915..7369 /gene="nsk1" /note="SPAC24H6.04, len:484aa; pfam match to entry PF00349 hexokinase, Hexokinases"
gene	/codon_start=1 /product="hexokinase" /protein_id="CAA90848.1" /db_xref="PID:g984701" /db_xref="GI:g984701" /translation="MSLIDAYHWPSRPISRKSGSINKLKTKTQDHLDELEDFPIPTPL LARVDRVSELYKGLTTNPGDVPPWPTWIIGTDGNEHSSTALDDGTMLRVCAYE VOGNKFITDSIKYRLPOLKLVGTREALFDYLADCKKFEVEVRGKSOMLEIETFS YPCVORSINDSLAVMTGFDPIDVGEGESVPLLSAALKRVGCNNVRNLMLSTGTCT LVASVPAGEFIEIVFTGCNACYIEKFSPELKHKKDFEDNMNMLINCDCPFGNO HVLYRTGYDYADVIDEESPGRGIGOTYEKMIAAGCYGIDLRILLDTBOGALFNQDVT KRDLADIATSYLSAEVDPEMNDDETCTEEFYGLKTTEENOFIRACELIGTRS ARLSGCGICALVKRNKRSMTVGTDGSYTINLPKRDLAQAFLDIIEBEGSRVVIT PAEDSSGGAAVLASLEAKRAFTSDIIAEHLKN"
misc_feature	6377..6454 /gene="nsk1" /note="PSO0378 Hexokinases signature"
gene	9276..11066 /gene="cdc25" /gene="cdc25" 9276..11066 /gene="cdc25" /note="SPAC24H6.05, len:596aa; conflict with published sequence: Pfam match to entry PF00581 Rhodanese, Rhodanese-like domain"
CDS	/product="H-phase inducer phosphatase" /protein_id="CAA90849.1" /db_xref="PID:g984702"

Query Match	83.28;	Score 15.8;	DB 8;	Length 36833;
Best Local Similarity	89.58;	Pred. No. 1.6e+02;		
Matches 17;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;

```

QY      1 AGGCTTCTTAGCCATCA 19
        | ||| ||||| |||||
Dio     42 ACGCTATCTTAGCCATCA 24

```

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:35 ; Search time 255.05 Seconds

(without alignments)
18.638 Million cell updates/sec

Title: US-09-325-095-25

Sequence: 1 AGCGTTCTTTAGCCATCA 19

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	3412	1 051156	Human p110 CDNA. R
2	19	100.0	3207	1 051155	p110 CDNA. Recombi
3	19	100.0	3488	1 057012	Pdins 3-kinase 11
4	15.4	81.1	2880	1 X14318	H. pylori GHPO 127
5	15	78.9	8196	1 V09699	Porcine retrovirus
6	15	78.9	1974	1 V09703	Porcine retrovirus
7	15	78.9	8209	1 V09700	Porcine retrovirus
8	14.8	77.9	1325	1 036666	Potato tuber PPO c
9	14.8	77.9	1670	1 056132	Potato polyphenol
10	14.8	77.9	682	1 T78398	Potato polyphenol
11	14.8	77.9	2370	1 V48311	Human cell-volume
12	14.8	77.9	2346	1 V74190	Human cell-volume
13	14.8	77.9	3166	1 X14374	Human cell-volume
14	14.8	77.9	2085	1 X01362	Nucleobase permeas
15	14.8	77.9	1907	1 X20307	Borrelia burgdorferi
16	14.8	77.9	1924	1 V72243	G. max SBP2 CDNA.
17	14.4	75.8	17041	1 Q21065	Genomic DNA of hum
18	14.4	75.8	3980	1 062696	Sequence of human
19	14.4	75.8	17041	1 067057	PACBP38 DNA. PACA
20	14.4	75.8	6000	1 T06308	Arabidopsis pathog
21	14.4	75.8	7000	1 T37476	RPP5 downy mildew
22	14.4	75.8	2329	1 X14596	H. pylori GHPO 175
23	14.2	74.7	1757	1 Q49428	Cytochrome P450 ho
24	14.2	74.7	2652	1 051233	Sequence encoding
25	14.2	74.7	4008	1 064827	Leukotoxin Apptina
26	14.2	74.7	110000	1 T58840.4	Continuation (5 of
27	14.2	74.7	7721	1 T73420	ApptinABD gene. P
28	14.2	74.7	1757	1 T94666	Petunia flavonoid
29	14.2	74.7	2448	1 V01882	Human nel-related
30	14.2	74.7	3198	1 V01883	Human nel-related
31	14.2	74.7	409	1 V75559	Staphylococcus aur
32	14.2	74.7	1903	1 X14490	H. pylori GHPO 113
33	14.2	74.7	1227	1 X14478	H. pylori GHPO 108
34	14.2	74.7	1465	1 X14121	H. pylori GHPO 153
35	14.2	74.7	1262	1 X27327	Human secreted pro
36	14.2	74.7	8241	1 N81439	Factor VIII CDNA. 1
37	14.2	74.7	73.7	1 N81096	CDNA sequence enco
38	14.2	74.7	7059	1 N81543	Human factor VIII-
39	14.2	74.7	7440	1 N80689	Sequence encoding
40	14.2	74.7	7587	1 N50137	Zucchin ACC synth
41	14.2	74.7	15155	1 N50107	DNA encoding facto
42	14.2	74.7	3852	1 N50094	DNA encoding facto
43	14.2	74.7	8975	1 N50054	Human factor VIII

ALIGNMENTS

44 14 73.7 7440 1 N50375 DNA sequence encod
c 45 14 73.7 4464 1 X22901 MO 99/07855 SeqID

```

RESULT 1
051156/c
ID 051156 standard; CDNA; 3412 BP.
AC 051156:
DE 12-APR-1994 (first entry)
DE Human p110 CDNA.
KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ds.
OS Human.
FH Key
FT cas
FT Location/Qualifiers
FT 1..3207
FT /*tag= a
FT /note= "PI3- kinase p110"
MO9321328-A.
PD 28-OCT-1993.
PR 13-APR-1993; G00761.
PR 13-APR-1992; GB-008135.
PA (UDW-) LUDWIG INST CANCER RES.
PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI Parker PJ, Volinia S, Waterfield MD;
DR WP1: 93-351738/44.
DR P-PDB: R43342.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity; useful for controlling cell proliferation
PS Claim 7; Fig 16; 146bp; English.
CC Southern blot analysis was performed using a bovine CDNA probe contg
CC a fragment of a PI3-kinase-encoding sequence and human CDNA isolated
CC from a CDNA library constructed from mRNA isolated from the human
CC cell line K562. Positive clones were sequenced to give the human
CC PI3 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with PI3 kinase activity, and is useful for
CC screening for (ant)agonists of PI3 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also 051155 and Q57522-3.
SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 100.0%; Score 19; DB 1; Length 3412;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCGTTCTTTAGCCATCA 19
|||
DB 878 AGCGTTCTTTAGCCATCA 860

RESULT 2
051155/c
ID 051155 standard; CDNA; 3207 BP.
AC 051155:
DE 12-APR-1994 (first entry)
DE p110 CDNA.
KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ss.
FH Key
FT cas
FT Location/Qualifiers
FT 1..3207
FT /*tag= a

```

```

FM      /note- "pi3- kinase p110"
PM      WO932328-A.
PD      28-OCT-1993.
PR      13-APR-1993; G00761.
PR      13-APR-1992; GB-008135.
PA      (LUDM-) LUDWIG INST CANCER RES.
PI      Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI      ParkerPJ, Volinia S, Waterfield MD;
PI      WPI: 93-351738/44.
DR      P-PSDB; R43341.
PT      Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PS      activity, useful for controlling cell proliferation
CC      Claim 7; Fig 9; 146pp; English.
CC      Southern blot analysis was performed using a bovine cDNA probe contg.
CC      a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
CC      from a cDNA library constructed from mRNA isolated from the human
CC      cell line KG1a. Positive clones were sequenced to give the human
CC      PI3 kinase p110 sequence shown. This sequence has 95 percent
CC      homology with the bovine sequence. The domain encoding residues 19-
CC      100 of human p110 is sufficient to encode the kinase which will
CC      associate with the p85 kinase subunit. The gene may be used to
CC      provide a protein with PI3 kinase activity, and is useful for
CC      screening for (ant)agonists of PI3 kinase activity which could be
CC      useful for stimulation or inhibition of cell proliferation and hence
CC      prophylaxis or therapy. Platelet or neutrophil activity or blood
CC      glucose levels can be controlled using the kinase.
CC      See also Q51156, Q59012-23 and Q57522-3.
SQ      Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match      100.0%; Score 19; DB 1; Length 3207;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGCGTTCTTACCATCA 19
        |||||||
Db      878 AGCGTTCTTACCATCA 860

RESULT 3
O57012/c      057012 standard; cDNA to mRNA; 3498 BP.
AC      057012:
DE      31-AUG-1994 (first entry)
KW      Ptdins 3-kinase 110 KD catalytic subnit cDNA.
KW      110 KD catalytic subunit; phosphatidy1 inositol 3-kinase;
KW      transfection; schizosaccharomyces pombe; nmt promoter; thiamine;
KW      Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
KW      blood vessel plaques; ss.
OS      Bos taurus.
FH      Key
FT      cds
FT      Location/Qualifiers
        1..3207
        /*tag= a
        /product= p110

WO9403409-A.
PD      17-FEB-1994.
PR      05-AUG-1993; G01651.
PR      05-AUG-1992; GB-016654.
PA      (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PI      Goode NT, Nurse PM, Parker PJ, Waterfield MD;
PI      WPI: 94-065697/08.
DR      P-PSDB; R46294.
PT      Eukaryotic cells transformed with mammalian phospholipid or
PT      protein kinase DNA - useful in assays for compounds involved in
PT      cell growth regulation and for treating cancers
PS      Disclosure: Fig 1; 71pp; English.
CC      This sequence encodes the 110 KD catalytic subunit of the phosphatidy1
CC      inositol (Ptdins) 3-kinase. This sequence was transformed into
CC      Schizosaccharomyces pombe cells under the regulatory control of the
CC      nmt promoter in an embodiment of the invention. In the presence of
CC      thiamine the promoter is inactive and the cells carrying the Ptdins
CC      catalytic subunit plasmid grow as the parental strain. In the absence
CC      of thiamine the nmt promoter functions and the Ptdins 3-kinase

```

```

CC      catalytic subunit in induced. Ptdins activity is substantially
CC      increased under these conditions. Cells containing constructs such as
CC      this, are useful in assays for detecting compounds involved in cell
CC      growth regulation. It is also used as the basis for detecting
CC      compounds for treating cancers and the formation of blood vessel
CC      plaques.
SQ      Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match      100.0%; Score 19; DB 1; Length 3498;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGCGTTCTTACCATCA 19
        |||||||
Db      878 AGCGTTCTTACCATCA 860

RESULT 4
X14318/c      X14318 standard; DNA; 2880 BP.
ID      X14318:
AC      X14318:
DE      31-MAR-1999 (first entry)
DE      H. pylori GHPD 127 gene.
KW      GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW      peptic ulcer disease; ss.
OS      Helicobacter pylori.
FH      Key
FT      cds
FT      Location/Qualifiers
        16..2817
        /*tag= a

WO9843478-A1.
PD      08-OCT-1998.
PR      01-APR-1998; U06371.
PR      29-JUL-1997; US-902615.
PR      01-APR-1997; US-833457.
PR      24-JUN-1997; US-881227.
PA      (HUMR-) HUMAN GENOME SCI INC.
PA      (IMMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI      Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
PI      WPI: 98-542293/46.
DR      P-PSDB; W98599.
PT      New isolated Helicobacter polynucleotides - used to develop products
PT      for the diagnosis, prevention and treatment of Helicobacter
PT      infections and gastrointestinal diseases
PS      Claim 1; Page 1195-1199; 2034pp; English.
CC      This sequence represents a polynucleotide of the invention. It was
CC      isolated from Helicobacter pylori and encodes a H.pylori GHPD protein.
CC      The polypeptides can be used for preventing or treating Helicobacter
CC      infections, and gastroduodenal diseases associated with these
CC      infections, including acute, chronic, and atrophic gastritis, and peptic
CC      ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC      for the production of antibodies. The products can also be used for
CC      detection and diagnosis.
SQ      Sequence 2880 BP; 997 A; 553 C; 589 G; 741 T;

Query Match      81.1%; Score 15.4; DB 1; Length 2880;
Best Local Similarity 94.1%; Pred. No. 36;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 AGCGTTCTTACCAT 17
        |||||||
Db      2690 AGCGTTCTTATCCAT 2674

RESULT 5
WO9699/c      WO9699 standard; DNA; 8196 BP.
ID      WO9699:
AC      WO9699:
DE      19-MAY-1998 (first entry)
DE      Porcine retrovirus DNA encoding, GAG, POL and ENV.
KW      Porcine retrovirus; PORV; POL protein; ENV protein; GAG protein;
KW      vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds...

```



```

FT misc_feature 1173
FT /*tag- b
FT /label- site_E
FT /note- "insertion site"
FT misc_feature 1645
FT /*tag- c
FT /label- site_F
FT /note- "insertion site"
FT MO9403607-A.
PN 17-FEB-1994.
PF 27-JUL-1993. E01988.
PR 30-JUL-1992. EP-402199.
PA (KEYG-) KEYGENE NV.
PI Bachem C, Zabeau M.
DR WPI: 94-065696/08.
PT DNA constructs contg. antisense polyphenol oxidase genes - used
PT for producing plants, partic. potatoes with reduced
PT susceptibility to bruising and browning
PS Claim 11: Fig 1b; 30pp; English.
CC A potato tuber cDNA clone bank was made and screened with a probe
CC corresp. to the sequence for a leaf PPO gene. The isolated clones
CC clones from each class were PKG45-8 (056131) and PKG59-4 (056132).
CC The latter is thought to be uniquely expressed in tubers. Antisense
CC constructs were made using the two full-length clones. The
CC constructs were introduced into 2 different lines of potato via
CC Agrobacterium tumefaciens; in field trials, potatoes derived from
SO regenerants showed reduced bruising.
SQ Sequence 1670 BP; 480 A; 337 C; 373 G; 480 T;

Query Match 77.9%; Score 14.8; DB 1; Length 1670;
Best Local Similarity 88.9%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GGCCTTCTTAGCCATCA 19
DB 1058 GGTTCCTTTGCGCATCA 1075

RESULT 10
T78398
ID T78398 standard; DNA; 682 BP.
AC T78398;
DT 23-MAR-1998 (first entry)
DE Potato polyphenol oxidase GPO6 genomic DNA.
KM Polyphenol oxidase; PPO; browning; fruit; vegetable;
KW genomic DNA amplification; ds.
OS Solanum tuberosum.
PN WO9729193-A1.
PD 14-AUG-1997.
PE 24-JAN-1997. AU0041.
PR 16-SEP-1996. AU-002361.
PR 05-FEB-1996. AU-007856.
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
PI Robinson SP.
DR WPI: 97-415348/38.
DR P-PSDB: W23687.
PT Preparation of polyphenol oxidase - by genomic DNA
PT amplification, useful to control browning reactions in fruit and
PT vegetables
PS Claim 40: Fig 24; 53pp; English.
CC A method has been developed for preparing a nucleic acid sequence
CC encoding polyphenol oxidase (PPO), or a fragment or derivative. The
CC method comprises amplifying genomic DNA isolated from plant tissue with
CC sense and antisense primers corresponding to conserved PPO gene regions.
CC The present sequence encodes a specifically claimed polyphenol oxidase.
CC Sense nucleic acid sequences can be used to increase or, by
CC co-suppression, decrease PPO activity in plants, while antisense
CC nucleic acid sequences reduce activity. Control of PPO activity allows
CC browning reactions in fruit and vegetables to be controlled, while
CC avoiding the need for chemicals, e.g. sulphur dioxide. Many PPO genes
CC lack introns, and can therefore be amplified directly from genomic DNA.

```

```

CC eliminating the need for separation of RNA and synthesis of cDNA. Also,
CC only small samples are needed and fragment size can be predicted,
CC allowing bands of appropriate size to be selected for cloning.
SQ Sequence 682 BP; 186 A; 138 C; 158 G; 200 T;

Query Match 77.9%; Score 14.8; DB 1; Length 682;
Best Local Similarity 88.9%; Pred. No. 64;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GGCCTTCTTAGCCATCA 19
DB 557 GGTTCCTTTGCGCATCA 574

RESULT 11
V48311/C
ID V48311 standard; cDNA; 2370 BP.
AC V48311;
DT 16-NOV-1998 (first entry)
DE Human cell-volume regulating kinase h-sgk.
KW human; cell-volume; kinase; h-sgk; diabetes mellitus;
KW renal insufficiency; inflammation; Alzheimer's disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 43..138
FT /*tag- a
FT /product- "Kinase h-sgk"
FT EP-861896-A2.
PN 02-SEP-1998.
PD 27-JAN-1998; 101338.
PE 28-FEB-1997; DE-008173.
PA (DADE-) DADE BEHRING MARBURG GMBH.
PI Lang F, Waldegger S.
DR WPI: 98-449109/39.
DR P-PSDB: W77219.
DE New nucleic acid encoding cell-volume regulating kinase h-sgk and
DE related proteins - used for diagnosis and treatment of diseases
PT involving changes in cell volume, e.g. renal insufficiency,
PT inflammation, infections etc.
PS Disclosure: Fig 1; 15pp; German.
CC The human cell-volume regulating kinase h-sgk is inhibited by the
CC swelling of cells (or presence of urea), whereas cell shrinkage
CC stimulates its expression. The nucleic acid h-sgk, and fragments, are
CC particularly used to detect changes in cell volume, specifically for
CC diagnosis of conditions that involve such changes, e.g. hyper- and hypo-
CC natremia, diabetes mellitus, renal insufficiency, hypercatabolism,
CC hepatic encephalopathy, inflammation, microbial/viral infection, fructose
CC intolerance, hyper- and hypo-glycaemia and Alzheimer's disease.
CC The nucleic acid, protein and products including receptors that bind
CC h-sgk, can be used to treat these disorders.
SQ Sequence 2370 BP; 636 A; 517 C; 513 G; 704 T;

Query Match 77.9%; Score 14.8; DB 1; Length 2370;
Best Local Similarity 88.9%; Pred. No. 72;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AGCCTTCTTAGCCATC 18
DB 1994 AGCCTTCTTAGCCATC 1977

RESULT 12
V74190/C
ID V74190 standard; DNA; 2346 BP.
AC V74190;
DT 15-MAR-1993 (first entry)
DE Human sgk DNA.
KW Serum glucocorticoid regulated kinase; sgk; human; treatment; inhibitor;
KW serine/threonine protein kinase family; antagonist; diabetic nephropathy;
KW chronic renal failure; inflammation; Alzheimer's disease; wound; ss.
OS Homo sapiens.

```

FH Key Location/Qualifiers
 FT CDS 45..1340
 FT /*tag= a
 FT /product= "sgk"
 FT /transl_except= (pos:1185..1187, aa:asp)
 PN EP-887081-A2.
 PD 30-DEC-1998.
 PR 27-MAY-1998; 304189.
 PR 27-JUN-1997; US-051124.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Kumar JM;
 DR MPI; 99-047627/05.
 DR P-PSDB; W80139.
 PT Treating chronic renal failure, diabetic nephropathy and Alzheimer's
 PT disease - by administration of nucleic acids and antagonists which
 PT inhibit activity or expression of human serum glucocorticoid
 PT regulated kinase (sgk), a serine/threonine protein kinase
 PS Disclosure: Page 14-15; 17pp; English.
 CC This sequence encodes a novel human serum glucocorticoid regulated kinase
 CC (sgk) protein which is a member of the serine/threonine protein kinase
 CC family. This protein is used for the treatment of a subject having need
 CC to inhibit/antagonise activity or expression of human sgk polypeptide
 CC e.g. for the treatment of chronic renal failure, diabetic nephropathy,
 CC inflammation, Alzheimer's disease and wounds.
 SO Sequence 2346 BP; 608 A; 520 C; 518 G; 700 T;

Query Match 77.9%; Score 14.8; DB 1; Length 2346;
 Best Local Similarity 88.9%; Pred. No. 72;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTGGCATTG 18
 Db 1999 ATGCTTCTTGGCATTG 1982

RESULT 13
 ID X14274 standard; DNA: 3166 BP.
 AC X14274;
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHP0 1661 gene.
 KW GHP0 protein: Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease; ss.
 OS Helicobacter pylori.
 FH Key Location/Qualifiers
 FT CDS 51..3116
 FT /*tag= a
 PN W08843478-A1.
 PD 08-OCT-1998.
 PR 01-APR-1998; U06371.
 PR 29-JUL-1997; US-902615.
 PR 01-APR-1997; US-833457.
 PR 24-JUN-1997; US-881227.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (IMMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PI Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
 DR MPI; 98-342293/46.
 DR P-PSDB; W85555.
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastroduodenal diseases
 PS Claim 1; Page 1082-1087; 2054pp; English.
 CC This sequence represents a polynucleotide of the invention. It was
 CC isolated from Helicobacter pylori and encodes a H. pylori GHP0 protein.
 CC The polypeptides can be used for preventing or treating Helicobacter
 CC infections, and gastroduodenal diseases associated with these
 CC infections, including acute, chronic, and atrophic gastritis, and peptic
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
 CC for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 SO Sequence 3166 BP; 1146 A; 448 C; 622 G; 950 T;

Query Match 77.9%; Score 14.8; DB 1; Length 3166;
 Best Local Similarity 88.9%; Pred. No. 74;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTGGCATTG 18
 Db 265 AGGCTATCTTGGCATTG 282

RESULT 14

ID X01362 standard; cDNA: 2085 BP.
 AC X01362;
 DT 16-APR-1999 (first entry)
 DE Nucleobase permease Ysp11 coding sequence.
 KW Ysp11; nucleobase permease; nucleobase binding affinity;
 KW drug screening assay; abnormal physiology; abnormal development;
 KW abnormal proliferation disease; cancerous condition;
 KW degenerative condition; ss.
 OS Mus sp.
 FH Key Location/Qualifiers
 FT CDS 48..1883
 FT /*tag= a
 FT /product= Ysp11
 FT 366..1883
 FT /note= "Ysp11 form 2 created by alternative splicing"
 FT 1515..1883
 FT /note= "Ysp11 form 3 created by alternative splicing"
 FT 1659..1883
 FT /note= "Ysp11 form 4 created by alternative splicing"
 FT /*tag= d

US5858707-A.
 PN 12-JAN-1999.
 PD 03-JUL-1996; 677049.
 PR 03-JUL-1995; US-000788.
 PR 03-JUL-1996; US-677049.
 PA (SCHE) SCHERING CORP.
 PI Bazan JF, Guimaraes MJ, McLanahan TK, Zlotnik A;
 DR MPI; 99-152103/13.
 DR P-PSDB; W73924.
 PT DNA encoding nucleobase permease polypeptides - useful for producing
 PT recombinant polypeptides
 PS Claim 10; Fig 2; 49pp; English.
 CC This sequence encodes the nucleobase permease of the invention.
 CC designated Ysp11. The nucleic acids are useful for producing recombinant
 CC polypeptides that have nucleobase binding affinity or nucleobase permease
 CC activity, for use in drug screening assays. The nucleobase permease can
 CC also be used in the treatment of conditions associated with abnormal
 CC physiology or development, including abnormal proliferation diseases such
 CC as cancerous conditions or degenerative conditions.
 SO Sequence 2085 BP; 364 A; 623 C; 548 G; 550 T;

Query Match 77.9%; Score 14.8; DB 1; Length 2085;
 Best Local Similarity 88.9%; Pred. No. 71;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGCTTCTTGGCATTG 19
 Db 1699 GGCTTCTTGGCATTG 1716

RESULT 15

ID X20307 standard; DNA: 1907 BP.
 AC X20307;
 DT 04-MAY-1999 (first entry)
 DE Borrelia burgdorferi polynucleotide sequence #60.
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;

KW Infection; diagnosis; characterisation; detection; ds.
 US Borrelia burgdorferi.
 PN W09858943-A1.
 PD 30-DEC-1998.
 PF 18-JUN-1998; 012764.
 PR 03-SEP-1997; US-057483.
 PR 20-JUN-1997; US-050359.
 PR 22-JUL-1997; US-053344.
 PR 22-JUL-1997; US-053377.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMUNE INC.
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
 PI White OR:
 DR WPI; 99-081217/07.
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 PS Claim 1: Page 1042-1043; 1128pp; English.
 CC X20248 to X20402 represent polynucleotide sequences isolated from
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 SQ Sequence 1907 BP; 615 A; 289 C; 164 G; 834 T;

Query Match 77.9%; Score 14.8; DB 1; Length 1907;
 Best Local Similarity 88.9%; Pred. No. 70;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GGCTTCTTAGCCATCA 19
 ||||||| ||||| |
 DB 1690 GGCTTCTTAGCCATCA 1707

Search completed: September 13, 1999, 15:59:36
 Job time: 4803 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:56; Search time 1694.61 Seconds

(without alignments)
22.116 Million cell updates/sec

Title: US-09-325-095-25

Perfect score: 19

Sequence: 1 AGGCTTCTTATGCCATCA 19

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database:

EST.*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	89.5	444	26	W39661	W39661 zc19g05.r1
2	17	89.5	343	39	AA860212	AA860212 ak48d05.s
3	17	89.5	452	47	A1493815	A1493815 q294e09.x
4	17	89.5	422	47	A1493925	A1493925 q294f09.x
5	16	84.2	227	37	AA667104	AA667104 vq88f07.r
6	15.8	83.2	717	22	R81085	R81085 EST22e01.WA
7	15.8	83.2	320	23	H34794	H34794 EST112038.R
8	15.8	83.2	603	25	W12707	W12707 ma68d08.r1
9	15.8	83.2	576	31	AA277365	AA277365 var8h10.r
10	15.8	83.2	470	33	AA445601	AA445601 v158a12.r
11	15.8	83.2	464	34	AA462201	AA462201 vg73b01.r
12	15.8	83.2	514	34	AA466787	AA466787 vd90a12.r
13	15.8	83.2	502	36	AA641040	AA641040 nr72d10.s
14	15.8	83.2	331	42	A1154471	A1154471 ua02h12.r
15	15.8	83.2	521	43	A1101017	A1101017 EST210306
16	15.8	83.2	438	45	A1010636	A1010636 EST205087
17	15.8	83.2	446	45	A1327456	A1327456 ma68d08.y
18	15.8	83.2	452	45	A1373771	A1373771 q254f03.x
19	15.8	83.2	475	46	AA963348	AA963348 UI-R-E1-g
20	15.8	83.2	280	48	AT548121	AT548121 UI-R-C3-s
21	15.8	83.2	518	49	A1639435	A1639435 rx04153s
22	15.8	83.2	216	50	AA031440	AA031440 AV031440
23	15.8	81.1	371	31	AA290151	AA290151 vb35b02.r
24	15.4	81.1	277	39	AA834058	AA834058 cf25g04.s
25	15.4	81.1	266	47	A1462605	A1462605 vb35b02.x
26	15.4	81.1	378	47	AT508508	AT508508 vb35b02.y
27	15	78.9	347	26	W61987	W61987 md84c02.r1
28	15	78.9	472	29	AA161835	AA161835 ms17d03.r
29	15	78.9	211	29	AA169006	AA169006 ms35g04.r
30	15	78.9	439	30	AA212613	AA212613 mu44f12.r
31	15	78.9	493	34	AA472626	AA472626 vhl1b04.r
32	15	78.9	429	39	AA879952	AA879952 vm02d04.r
33	15	78.9	349	42	A1152256	A1152256 ud21c01.r
34	15	78.9	382	44	AU043395	AU043395 AU043395
35	15	78.9	288	46	A1414297	A1414297 md84c01.x
36	15	78.9	175	46	A1414750	A1414750 ma88e08.x
37	15	78.9	349	46	A1415238	A1415238 md14b07.x
38	15	78.9	519	48	AT607060	AT607060 ms17d03.y
39	14.8	77.9	299	20	T03898	T03898 seg2417.4HB
40	14.8	77.9	407	21	R09422	R09422 y122h09.r1
41	14.8	77.9	296	22	H03079	H03079 y146h12.r1
42	14.8	77.9	429	22	R31727	R31727 yb63e11.r1
43	14.8	77.9	231	22	R72272	R72272 y188e12.s1
44	14.8	77.9	476	24	H67434	H67434 yd52e01.r1
45	14.8	77.9	533	49	AT652703	AT652703 wb30g07.x

ALIGNMENTS

RESULT 1
LOCUS W39661
DEFINITION zc19g05.r1 Soares parathyroid tumor NBHPA Homo sapiens cDNA clone
IMAGE:32808 5' similar to PIR:A35419 A35419 neutrophil protein -
ACCESSION W39661
NID 9131370

VERSION M39661.1 GI:1321370
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 444)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 TITLE The Washu-Merck EST Project
 JOURNAL Unpublished (1993)
 COMMENT On Nov 29, 1993 this sequence version replaced gi:430062.
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 705 Std Error: 0.00
 Seq primer: mod. REGA+ET
 High quality sequence stop: 302.
 Location/Qualifiers
 1..444
 /organism="Homo sapiens"
 /db_xref="GDB:1260466"
 /db_xref="taxon:9606"
 /clone="IMAGE:322808"
 /clone_1lb="Soares_parityroid_tumor_NbHPA"
 /tissue_type="parathyroid tumor"
 /dev_stage="adult"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: parathyroid gland; Vector: pT733
 (Pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer
 15'-
 TGTACCAATCGAAGTGGAGCGCGCCGCAATTTTCTTTTCTTTT
 T-3'), double-stranded cDNA was size selected, ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of a modified pT73
 vector (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid
 adenomas was kindly provided by Dr. Stephen Marx, National
 Institute of Diabetes and Digestive and Kidney Diseases,
 NIH."

BASE COUNT 161 a 67 c 131 g 83 t 2 others
 ORIGIN
 Query Match 89.58; Score 17; DB 26; Length 444;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 AGGCTTCTTAGCCAT 17
 ||||||||||||||||
 Db 392 AGGCTTCTTAGCCAT 408

RESULT 2
 LOCUS AA860212 343 bp mRNA EST 04-JAN-1999
 DEFINITION ak48605.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1409193
 3', mRNA sequence.
 ACCESSION AA860212
 NID 92954207
 VERSION AA860212.1 GI:2954207

KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 343)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1400930.
 CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-dio.llnl.gov/dbp/image/image.html
 Insert Length: 796 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 315.
 Location/Qualifiers
 1..343
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1409193"
 /clone_1lb="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT733-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer 15'-
 TGTACCAATCGAAGTGGAGCGCGCCGCAATTTTCTTTTCTTTT 3').
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 99 a 67 c 55 g 122 t
 ORIGIN
 Query Match 89.58; Score 17; DB 39; Length 343;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 AGGCTTCTTAGCCAT 17
 ||||||||||||||||
 Db 260 AGGCTTCTTAGCCAT 276

RESULT 3
 LOCUS A1493915 452 bp mRNA EST 18-MAR-1999
 DEFINITION g294609.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
 IMAGE:2042248 3', mRNA sequence.
 ACCESSION A1493915
 NID 94394918
 VERSION A1493915.1 GI:4394918
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 452)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3137579.

FEATURES
SOURCE
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 517 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 429.
Location/Qualifiers

1. 452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="14"
/clone_lib="Soares_pregnant_uterus_JbHPV"
/dev_stage="adult"
/sex="female"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dt) primer 15',
AACTGGAGAAATTCGCGCCCTTTTCTTTTCTTTT 3',
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 125 a 105 c 79 g 143 t
ORIGIN
Query Match 89.5%; Score 17; DB 47; Length 452;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGCTTCTTTAGCCAT 17
|||||
Db 252 AGGCTTCTTTAGCCAT 268
RESULT 4
A1493925 422 bp mRNA EST 18-MAR-1999
LOCUS q284f09.x1 Soares_pregnant_uterus_NbHPV Homo sapiens cDNA clone
DEFINITION IMAGE:2042249 3', mRNA sequence.
ACCESSION A1493925
NID 94394928
VERSION A1493925.1 GI:4394928
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 422)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
JOURNAL On May 18, 1998 this sequence version replaced gi:3137589.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 516 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 416.

FEATURES
SOURCE

Location/Qualifiers
1. 422
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2042249"
/clone_lib="Soares_pregnant_uterus_NbHPV"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dt) primer 15',
AACTGGAGAAATTCGCGCCCTTTTCTTTTCTTTT 3',
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 118 a 94 c 72 g 138 t
ORIGIN
Query Match 89.5%; Score 17; DB 47; Length 422;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGCTTCTTTAGCCAT 17
|||||
Db 254 AGGCTTCTTTAGCCAT 270
RESULT 5
AA667104 227 bp mRNA EST 19-NOV-1997
LOCUS v988f07.r1 Knowles Solter mouse blastocyst B3 Mus musculus cDNA
DEFINITION clone IMAGE:1109413 5' similar to gb:M8194 Mouse fibronectin
(MOUSE);, mRNA sequence.
ACCESSION AA667104
NID 92625805
VERSION AA667104.1 GI:2625805
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 227)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theisinger, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
JOURNAL On May 9, 1995 this sequence version replaced gi:803003.
COMMENT

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:607581
Trace considered overall poor quality
High quality sequence stop: 1.

FEATURES
SOURCE

1. 227
Location/Qualifiers
/organism="Mus musculus"
/strain="B6D2 F1/7"
/db_xref="taxon:10090"
/clone="IMAGE:1109413"

```

/clone_lib="Knowles Solter mouse blastocyst B3"
/issue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pSPORT; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(dT):
5'-CGGTGACCGTCGACCGTTTATTTT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."
BASE COUNT      59 a      54 c      38 g      76 t
ORIGIN
Query Match      84.2%; Score 16; DB 37; Length 227;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 GCCTTCTTTAGCCATC 18
      |||||||
Db      12 GCCTTCTTTAGCCATC 27

RESULT 6
R81085/c      R81085      717 bp      mRNA      EST      12-JUN-1996
LOCUS      EST222e01 WATM1 Homo sapiens cDNA clone 22e01, mRNA sequence.
DEFINITION      R81085
ACCESSION      R81085
NID      9857688
VERSION      R81085.1 GI:857688
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      Boulland, F.
TITLE      1 (bases 1 to 717)
JOURNAL      Study of expressed sequences tags in adipose tissue 1995
COMMENT      Unpublished (1995)
*on May 8, 1995 this sequence version replaced gi:801224.

Contract: Frederic Boulland
Centre de Recherche sur l'Endocrinologie moleculaire et le
Developpement
CNRS
9, Rue Jules Hetzel, Meudon Bellevue, 92190 France
Tel: 33 1 45 07 52 87
Fax: 33 1 45 07 58 90
Email: boulland@infobiogen.fr
automatic cycle sequencing of PCR amplified insert
Insert Length: 851 Std Error: 0.00
High quality sequence stop: 290.
Location/Qualifiers
1. 717
/organism="Homo sapiens"
/strain="Caucasian"
/db_xref="taxon:9606"
/clone="22e01"
/notes="Vector: lambda gtl1; Site: 1: EcoRI; White adipose
tissue, sub cutaneous, adult, female. Purification of
polya mRNA, first strand priming with random
oligonucleotides. Ligation of adaptors EcoRI/NotI, cloning
in EcoRI site of lambda gtl1."
BASE COUNT      196 a      172 c      192 g      157 t
ORIGIN
Query Match      83.2%; Score 15.8; DB 22; Length 717;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 AGGCTTCTTTAGCCATCA 19
      |||||||
Db      252 AGGCTTCTTTAGCCATCA 234

RESULT 7
H34794/c      H34794      320 bp      mRNA      EST      13-MAR-1998
LOCUS      EST112038 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. cDNA 5'
DEFINITION      end, mRNA sequence.
ACCESSION      H34794
NID      9980211
VERSION      H34794.1 GI:980211
KEYWORDS      EST.
SOURCE      Rattus sp.
ORGANISM      Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE      1 (bases 1 to 320)
AUTHORS      Lee, N.H., Weinstein, K.G., Kirkness, E.F., Earle-Hughes, J.A.,
Fuldner, R.A., Marmaras, S., Glodek, A., Goodyne, J.D., Adams, M.D.,
Kerlavage, A.R., Fraser, C.M. and Venter, J.C.
TITLE      Comparative expressed-sequence-tag analysis of differential gene
expression profiles in PC-12 cells before and after nerve growth
factor treatment
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
MEDLINE      95396786
COMMENT      Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
For clone availability please contact the TIGR Database
(tdbinfo@tdb.tigr.org)
Insert Length: 855 Std Error: 0.00
Seq primer: M13 Reverse
High quality sequence stop: 1.
Location/Qualifiers
1. 320
/organism="Rattus sp."
/db_xref="ATCC (inhost):2066129"
/db_xref="taxon:10118"
/clone_lib="Rat PC-12 cells, NGF-treated (9 days)"
/notes="Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2:
XhoI; poly(A)+ RNA was purified from 9-day NGF treated
PC12 cells. cDNA was constructed using an oligo-dT primer
and directionally cloned using the Lambda Zap II Vector
Kit by Stratagene."
BASE COUNT      106 a      61 c      74 g      79 t
ORIGIN
Query Match      83.2%; Score 15.8; DB 23; Length 320;
Best Local Similarity 89.5%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 AGGCTTCTTTAGCCATCA 19
      |||||||
Db      22 AGGCTTCTTTAGCCATCA 4

RESULT 8
W12707      W12707      603 bp      mRNA      EST      02-OCT-1997
LOCUS      ma68d08.r1 Soares mouse p3NM919.5 Mus musculus cDNA clone
DEFINITION      IMAGE:315855 5' similar to SW:PEPI_RAT P07734 BRAIN SPECIFIC
POLYPEPTIDE pep-19 ;, mRNA sequence.
ACCESSION      W12707
NID      91286817
VERSION      W12707.1 GI:1286817

```


KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 603)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 21, 1992 this sequence version replaced g1:279349.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI: 206471
Seq primer: mob.REGA+ET
High quality sequence stop: 459.
Location/Qualifiers
1. 603
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:315855"
/clone_id="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5, TGTTACCAATCTGAGTGGAGCGCGCCGCAATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

BASE COUNT 156 a 165 c 145 g 137 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 25; Length 603;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTACCATCA 19
|||||
DB 299 AGGCTCTCTAAGCCATCA 317

RESULT 9
LOCUS AA277365 576 bp mRNA EST 01-APR-1997
DEFINITION v278h10.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:737539 5' similar to SW:PE19_HUMAN P48539 BRAIN SPECIFIC POLYPEPTIDE PEP-19.
ACCESSION AA277365
NID 91917845
VERSION AA277365.1 GI:1917845
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 470)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project

REFERENCE 1 (bases 1 to 576)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Dec 30, 1996 this sequence version replaced g1:1529530.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:454587
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 497.
Location/Qualifiers
1. 576
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:737539"
/clone_id="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5, TGTTACCAATCTGAGTGGAGCGCGCCGCAATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 138 a 194 c 136 g 108 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 31; Length 576;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTACCATCA 19
|||||
DB 474 AGGCTCTCTAAGCCATCA 492

RESULT 10
LOCUS AA445601 470 bp mRNA EST 03-JUN-1997
DEFINITION v158a12.r1 Soares mouse NDMH Mus musculus cDNA clone IMAGE:847966 5' similar to SW:PE19_HUMAN P48539 BRAIN SPECIFIC POLYPEPTIDE PEP-19. mRNA sequence.
ACCESSION AA445601
NID 92158284
VERSION AA445601.1 GI:2158284
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 470)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)
COMMENT On May 9, 1995 this sequence version replaced g1:802535.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:500118
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 466.
Location/Qualifiers

FEATURES

source

1. 470
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:847966"
/clone_lib="Soares mouse NBMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'.
TGTTACCAATCTGAGTGGAGCGCGCGAAGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p7T3 vector. RNA
provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

BASE COUNT 128 a 127 c 115 g 100 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 33; Length 470;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTTAGCCATCA 19
||||| ||||| ||||| |||||
DB 306 AGGCTCTTTAGCCATCA 324

RESULT 11
LOCUS AA462201 464 bp mRNA EST 10-JUN-1997
DEFINITION v973b01.r1 Soares mouse NBMH Mus musculus cDNA clone IMAGE:871561
ACCESSION AA462201
VERSION 92187092
KEYWORDS AA462201.1 GI:2187092
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 464)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced g1:1400786.

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:511041
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 422.
Location/Qualifiers

FEATURES

source

1. 464
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:871561"
/clone_lib="Soares mouse NBMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'.
TGTTACCAATCTGAGTGGAGCGCGCGAAGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p7T3 vector. RNA
provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

BASE COUNT 110 a 126 c 112 g 116 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 34; Length 464;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTTAGCCATCA 19
||||| ||||| ||||| |||||
DB 73 AGGCTCTTTAGCCATCA 91

RESULT 12
LOCUS AA466787 514 bp mRNA EST 11-JUN-1997
DEFINITION v990a12.r1 Soares mouse NBMH Mus musculus cDNA clone IMAGE:807838
ACCESSION AA466787
VERSION 92192927
KEYWORDS AA466787.1 GI:2192927
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 514)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 25, 1995 this sequence version replaced g1:637840.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:484182

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 510.

FEATURES

Source

Location/Qualifiers

1..514

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:807838"

/clone_1lb="Soares mouse NDMH"

/sex="male"

/tissue_type="heart"

/dev_stage="4 weeks"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAGAGGAGCGCGCGAAGTTTCTTTTCTTTTCTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not

I and Eco RI sites of the modified pT73 vector. RNA

provided by Dr. Minoru Ko, Wayne State Univ. Library

constructed and normalized by Bento Soares and M.Patima

Bonaldo."

BASE COUNT 117 a 155 c 110 g 132 t

ORIGIN

Query Match 83.2%; Score 15.8; DB 34; Length 514;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AGGCTTCTTTAGGCATCA 19

Db 149 AGGCTTCTTTAGGCATCA 167

RESULT 13
AA641040/c 502 bp mRNA EST 27-OCT-1997
LOCUS nt72d10.s1 NCI-CGAP-Pr24 Homo sapiens cDNA clone IMAGE:1173523 3'
DEFINITION similar to gb:D13748 EUKARYOTIC INITIATION FACTOR 4A-1 (HUMAN);,
mRNA sequence.

ACCESSION AA641040

NID 92566290

VERSION AA641040.1 GI:2566290

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 502)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)

JOURNAL On Apr 14, 1993 this sequence version replaced gi:692623.

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright,

Ph.D.

CDNA Library Preparation: Stratagene, Inc.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 146.
Location/Qualifiers

FEATURES

Source

1..502

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1173523"

/clone_1lb="NCI-CGAP-Pr24"

/tissue_type="invasive tumor (cell line)"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: prostate; Vector: Bluescript SK-; Site_1:

EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:

Oligo dT. Invasive prostate tumor cell line (HPV

immortalized). 5' adaptor sequence: 5' GAATTCGGCAGCAG 3'

3' adaptor sequence: 5' CTCGAGCTTTTCTTTTCTTTT 3'

Average insert size: 1.0 Kb."

BASE COUNT 120 a 125 c 120 g 137 t

ORIGIN

Query Match 83.2%; Score 15.8; DB 36; Length 502;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AGGCTTCTTTAGGCATCA 19

Db 49 AGGCTTCTTTAGGCATCA 31

RESULT 14
A1154471 331 bp mRNA EST 30-SEP-1998
LOCUS us02h12.r1 Soares 2NBMT Mus musculus cDNA clone IMAGE:1345607 5',
DEFINITION mRNA sequence.

ACCESSION A1154471

NID 93682940

VERSION A1154471.1 GI:3682940

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 331)

Marras,M., Haller,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucabara,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

JOURNAL On Nov 29, 1993 this sequence version replaced gi:636069.

COMMENT

Contact: Marras M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:694399

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 207.

Location/Qualifiers

1..331

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/map="3"

/clone="IMAGE:1345607"

/clone_1lb="Soares 2NBMT"

```

/sex="male"
/tissue_type="thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5';
1GTTCCATGTCGAGTGGAGCGCGGTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M. Fatima Bonaldo."

```

BASE COUNT

```

87 a 69 c 57 g 118 t

```

ORIGIN

```

Query Match      83.2%; Score 15.8; DB 42; Length 331;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 AGGCTTCTTTAGCCATCA 19
      11111111111111111111
DB      47 AGTTCTCTTAGCCATCA 65

```

RESULT 15

```

A1101017      521 bp      mRNA      EST      31-JAN-1999
LOCUS      EST210306 Normalized rat brain. Bento Soares Rattus sp. cDNA clone
DEFINITION      RBRBE25 3' end, mRNA sequence.
ACCESSION      A1101017
NID      93706015
VERSION      A1101017.1 GI:3706015
KEYWORDS      EST.
SOURCE      Rattus sp.
ORGANISM      Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```

```

REFERENCE      Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
AUTHORS      Kerlavage,A.R. and Adams,M.D.
TITLE      Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
Gene Index
JOURNAL      Unpublished (1998)
COMMENT      On Aug 21, 1998 this sequence version replaced.

```

Contact: Lee, NH

ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M1321.

FEATURES

```

source
1..521
  /organism="Rattus sp."
  /db_xref="ATCC (Inhost):2023897"
  /db_xref="taxon:10118"
  /clone="RBRBE25"
  /clone_11b="Normalized rat brain, Bento Soares"
  /note="Organ: brain; Vector: p773D-Pac; Site_1: EcoRI;
  Site_2: NotI"

```

```

BASE COUNT      130 a 132 c 136 g 123 t
ORIGIN

```

```

Query Match      83.2%; Score 15.8; DB 43; Length 521;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 AGGCTTCTTTAGCCATCA 19
      11111111111111111111
DB      450 AGGCTCTTTAGCCATCA 432

```

Search completed: September 13, 1999, 15:45:57
Job time: 4045 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:58 ; Search time 539.84 Seconds

(without alignments)
117.824 Million cell updates/sec

Title: US-09-325-095-27

Perfect score: 1 GGNATGATTCRCNCARGA 20

Scoring table:

Searched: 679419 seqs, 1590154680 residues

Database:

GenEmbl: *
1: gb_dal: *
2: gb_da2: *
3: gb_cm: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pr1: *
10: gb_pr2: *
11: gb_pr3: *
12: gb_ro: *
13: gb_st: *
14: gb_sts: *
15: gb_sy: *
16: gb_un: *
17: gb_vt: *
18: em_fun: *
19: em_htg: *
20: em_hum1: *
21: em_hum2: *
22: em_in: *
23: em_com: *
24: em_ov: *
25: em_ov: *
26: em_pat: *
27: em_ph: *
28: em_pl: *
29: em_to: *
30: em_sts: *
31: em_sy: *
32: em_un: *
33: em_vt: *
34: gb_htg1: *
35: gb_htg2: *
36: gb_in1: *
37: gb_in2: *
38: em_dal: *
39: em_da2: *
40: em_hum3: *
41: em_hum4: *
42: gb_pt4: *

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	16	80.0	3207	3	BOVPHOS3KN	M93252 Bovine phop

2	16	80.0	3808	3	SSP120	V10743 S. scrofa mR
3	16	80.0	3452	4	AF001076	AF001076 Gallus ga
4	16	80.0	3944	4	GU095088	GU095088 Gallus gall
5	16	80.0	3498	5	A37232	A37232 Sequence 1
6	16	80.0	4134	5	A50263	A50263 Sequence 1
7	16	80.0	4137	5	A50265	A50265 Sequence 3
8	16	80.0	339	8	CEU97661	CEU97661
9	16	80.0	2200	8	CEU97662	CEU97662
10	16	80.0	2634	8	SPU32583	SPU32583
11	16	80.0	5397	10	HSP13K	HSP13K
12	16	80.0	3424	10	HSP13K	HSP13K
13	16	80.0	2990	10	HSP13K	HSP13K
14	16	80.0	3207	10	HSU79143	HSU79143
15	16	80.0	3213	10	S67334	S67334
16	16	80.0	3207	12	MMU03279	MMU03279
17	16	80.0	3213	12	RNO012482	RNO012482
18	16	80.0	2752	12	RNO6710	RNO6710
19	16	80.0	3389	12	AF001075	AF001075
20	16	80.0	135259	34	AC005018	AC005018
21	16	80.0	17370	36	CEB0025	CEB0025
22	16	80.0	2837	36	CEBPS34	CEBPS34
23	16	80.0	5577	36	DDU23477	DDU23477
24	16	80.0	3165	36	DMP13K59F	DMP13K59F
25	16	80.0	188288	37	AC005639	AC005639
26	15	75.0	115602	9	HS1118D24	HS1118D24
27	14.8	74.0	2612	1	CGDNAROP	CGDNAROP
28	14.8	74.0	24800	1	SC9B1	SC9B1
29	14.8	74.0	11925	2	AE001013	AE001013
30	14.8	74.0	3331	3	BOVNTOMX	BOVNTOMX
31	14.8	74.0	3859	3	BTU88531	BTU88531
32	14.8	74.0	1997	4	AF001413	AF001413
33	14.8	74.0	1612	5	AS8887	AS8887
34	14.8	74.0	2987	7	SOYPI3K5X	SOYPI3K5X
35	14.8	74.0	2760	7	U10669	U10669
36	14.8	74.0	120787	8	AC002292	AC002292
37	14.8	74.0	3471	9	AB005910	AB005910
38	14.8	74.0	3261	9	HSR011121	HSR011121
39	14.8	74.0	3599	9	HSR011122	HSR011122
40	14.8	74.0	3599	9	HSR011123	HSR011123
41	14.8	74.0	4880	9	HSC2P13K1	HSC2P13K1
42	14.8	74.0	7654	10	HSC2P13K	HSC2P13K
43	14.8	74.0	3868	10	HSP110DEL	HSP110DEL
44	14.8	74.0	5296	10	HSP13K1NA	HSP13K1NA
45	14.8	74.0	62980	35	AC007771	AC007771

ALIGNMENTS

RESULT	1	BOVPHOS3KN	3207 bp	MAN	19-AUG-1992
LOCUS	1	BOVPHOS3KN	3207 bp	MAN	19-AUG-1992
DEFINITION	1	Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete			
ACCESSION		M93252			
NTD		9163519			
VERSION		M93252.1	GI:163519		
KEYWORDS		Phosphatidylinositol 3-kinase.			
SOURCE		Bos taurus cDNA to mRNA.			
ORGANISM		Bos taurus			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.			
AUTHORS		Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Fhan, R., Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F., Hsuan, J.J., Courtneidge, S.A., Parker, P.J. and Waterfield, M.D.			
TITLE		Phosphatidylinositol 3-kinase: Structure and expression of the 110 kD catalytic subunit			
JOURNAL		Cell 70, 419-429 (1992)			
MEDLINE		92354059			
FEATURES		Location/Qualifiers			
source		1..3207			

Scriptos Research Institute, 10550 N. Torrey Pines Road, Mall-drop
BCC-239, La Jolla, CA 92037, USA

FEATURES

source

Location/Qualifiers

1..3452

/organism="Gallus gallus"

/db_xref="taxon:9031"

/tissue_type="brain"

1..150

151..3354

/gene="c-p3k"

/note="p3k proto-oncogene"

151..3357

/note="c-p3k"

/codon_start=1

/product="phosphoinositide 3-kinase catalytic subunit"

/protein_id="AAB62534.1"

/db_xref="pid:g2245506"

/db_xref="GI:2245506"

/translation="MPPRSSGELMGTHLMPRLVCLPNCNITVLECLREATLT

IKHELEKARKYPLHQLDDESSYIFSVQAEKREFFETRRLCDRLFQPLKVI

EPVGNREKILNREIGFAGMPICEFMDVDEYDERRNINLVCKEAVDLINDSPH

SRALYCPPNVSSPELPHKIIYKLDKQIIVYIIVVSPNNKOKYTLKINHCVE

QVIAEAIKRTKRSMLSSDQLCVLEKQYILKVGCEYILEKPLSQYKIRSC

IMGRMNLIMAKESLYTLPDFTMPSSRSISATPTMGEAKAKSLMTNSL

RIRLCATYVNNIRIDIKIVRTGITHGSEPLCDNVNTPVCSNRMNWSLDMT

IPDLPRARCLSTCSVGRKGAKEHCPLAMGIMNFDYTDLVSGKALNLNAVH

GLEDLNPIGVTSNPKETPCLEEDWFSNPKEFDMYIEEHAWTISREGFY

SYAGSNRIARDNELRESDEKQLAICTRDLSEITQEKDFLESHHVCNTEIIP

KLLSVKNSRDEVAQMYCLYQMPRIKPKDQAMLLCNQPDVRAFAVRCLEKVT

DDKLSOYLIDVLYKYEYLDNOLVFLKALITNORIGHFFVHKLSEKHNKTVQ

RGLLESTCRACGMYLKLHROYEAMEKLINTDILNKEKDETQVOMFLVEOK

RPPMDALQGFISPLNPAHQNLRLSECRIMSSAKPLMLMNPIMSELVONNE

IIFKNGDQLKODMLTQIIRIMENIMONQGLDMLYGCISIDCYGLLEVSSHT

IMQIOCKGLKALQFNSHTLQMLKDKNKGEMDALIDFTSCAGCYATFLIG

DRHNSNIMVKDDQLFHDHGFLEHKKKRGYRREVPFLVQDPLIVISKGQECT

KTRERFQEMKYKAYLAIRQHANLFINLSMKSGSMPELQSFDDIAYIRKITALK

TEQDALEPKQMDAHHGWTMTMDIFHTIKQHAUN"

3358..3452

BASE COUNT 1073 a 672 c 783 g 924 t

ORIGIN

Query Match 80.0%; Score 16; DB 4; Length 3452;

Best Local Similarity 65.0%; Pred. No. 1e+02;

Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 1 GGNAGYGYATTCGNCARGA 20

|||:|:|:|:|:|:|:|:|:|

2560 GGACATGACTGCGTCAGCA 2579

RESULT 4

GGU95088 3944 bp DNA VRT 02-JAN-1999

LOCUS Gallus gallus progesterone receptor binding protein (rbp) gene,

complete cds.

ACCESSION U95088

NID 94100171

VERSION U95088.1 GI:4100171

KEYWORDS

SOURCE chicken.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;

Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

1 (bases 1 to 3944)

Sandhu, N.P. and Spelsberg, T.C.

Cloning and analysis of the gene encoding the chicken oviduct

progesterone receptor binding factor

Unpublished

2 (bases 1 to 3944)

Sandhu, N.P. and Spelsberg, T.C.

Direct Submission

Clinic, 200 First Street, S.W., Rochester, MN 55905, USA

FEATURES

source

Location/Qualifiers

1..3944

/organism="Gallus gallus"

/db_xref="taxon:9031"

/tissue_type="oviduct"

join(1711..1950,2076..2130,2360..2458,3589..3703)

/gene="rbp"

/product="progesterone receptor binding protein"

1711..3703

/gene="rbf"

join(1915..1950,2076..2130,2360..2458,3589..3614)

/gene="rbf"

/codon_start=1

/product="progesterone receptor binding protein"

/protein_id="AAD00767.1"

/db_xref="pid:g4100172"

/db_xref="GI:4100172"

/translation="MPPVQSPILKFTNYSALLVGNITGRKRYDYLPKPAEERIE

AEKKRELEERAKELAEASEDSILK"

3698..3703

BASE COUNT 1074 a 880 c 885 g 1155 t

ORIGIN

polya__signal

Query Match 80.0%; Score 16; DB 4; Length 3944;

Best Local Similarity 65.0%; Pred. No. 1e+02;

Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 1 GGNAGYGYATTCGNCARGA 20

|||:|:|:|:|:|:|:|:|:|

2102 GGGGATGATCAGCGCAGCA 2121

RESULT 5

A37232 3498 bp DNA PAT 05-MAR-1997

LOCUS Sequence 1 from Patent WO9403609.

DEFINITION A37232

ACCESSION 92294345

NID A37232.1 GI:2294345

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 3498)

Parker, J.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.

EUCARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE

OR PROTEIN KINASE AND ASSAYS USING THEM

Patent: WO 9403609-A 1 17-FEB-1994;

IMP CANCER RES TECH (GB)

Other publication JP 8503124T 960409.

Location/Qualifiers

1..3498

/organism="unidentified"

/db_xref="taxon:32644"

1..3207

/note="unidentified protein product"

/codon_start=1

/product="progesterone receptor binding protein"

/protein_id="CAA02284.1"

/db_xref="pid:g304808"

/db_xref="GI:2294346"

/translation="MPPRSSGELMGTHLMPRLVCLPNCNITVLECLREATLT

IKHELEKARKYPLHQLDDESSYIFSVQAEKREFFETRRLCDRLFQPLKVI

EPVGNREKILNREIGFAGMPICEFMDVDEYDERRNINLVCKEAVDLINDSPH

SRALYCPPNVSSPELPHKIIYKLDKQIIVYIIVVSPNNKOKYTLKINHCVE

QVIAEAIKRTKRSMLSSDQLCVLEKQYILKVGCEYILEKPLSQYKIRSC

IMGRMNLIMAKESLYTLPDFTMPSSRSISATPTMGEAKAKSLMTNSL

RIRLCATYVNNIRIDIKIVRTGITHGSEPLCDNVNTPVCSNRMNWSLDMT

IPDLPRARCLSTCSVGRKGAKEHCPLAMGIMNFDYTDLVSGKALNLNAVH

SHAGLSNLANDEINRENDKQALRAICTRDPISSETDEKDFLASHREYCTITPILP
KLLSVKNSNDEVAQMYCIVADWPPKPEQAMELIDONTDPDPAVGRGAVRLELYLT
DDKLSQVLIQVLYLKYQYLDNLVRELKALNTNOIGHEFWHLKSEMNKTVSQ
RFGLLSEYCRAGMYLKLHNOVAMEKILMTDILKOEKDEFOKOMEFLVQMR
RPDEFDALOGFISPLNPAHOLGNLEECRISMSAKRPLMTNWEOPDIJMSLEHPCNR
IIFKNGDLRODMETQITRIMENTWOGDIDRLPYGCCSTIGCAGLIEVRSHNT
IMQICCKGGLGALDFNSHTLHQLNDKNKGITTYAALDIFRSCAGICVATFIIIG
DRHNSINAVKDGOLFHDGFHLDHKKKEGKREKREPVLTODFLIVISGADECT
KTRFERFOEMCYKAYLIROHANLFINLFSMLGSGMPELOSFDIAYIRKTTALDK
TEOEALFEYFMKOMDANHGGWTKMDWIFHTIKOHALN"

BASE COUNT 1147 a 626 c 722 g 1003 t
ORIGIN
Query Match 80.0%; Score 16; DB 5; Length 3498;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGNAGYATTCGNCARGA 20
|||:||||:||||:|
Db 2410 GGGATGATTCGCGCAGA 2429

RESULT 6
LOCUS A50263 4134 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 1 from Patent WO9612024.
ACCESSION A50263
NID 92303320
VERSION A50263.1 GI:2303320
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 4134)
Stoyanov, B., Hancock, T. and Wetzker, R.
TITLE CLONING, EXPRESSION AND CHARACTERISATION OF A NOVEL FORM OF
PHOSPHATIDYLINOSITOL-3-KINASE
JOURNAL Patent: WO 9612024-A 1 25-APR-1996;
MAX PLANCK GEBIETSSCHAFT (DE)
COMMENT Other publication DE 4445562 960404.
FEATURES
source 1. 4134
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 1127 a 999 c 998 g 1010 t
ORIGIN
Query Match 80.0%; Score 16; DB 5; Length 4134;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGNAGYATTCGNCARGA 20
|||:||||:||||:|
Db 2820 GGTGATGATTCGCGCAGA 2839

RESULT 7
LOCUS A50265 4137 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 3 from Patent WO9612024.
ACCESSION A50265
NID 92303321
VERSION A50265.1 GI:2303321
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 4137)
Stoyanov, B., Hancock, T. and Wetzker, R.

TITLE CLONING, EXPRESSION AND CHARACTERISATION OF A NOVEL FORM OF
PHOSPHATIDYLINOSITOL-3-KINASE
JOURNAL Patent: WO 9612024-A 3 25-APR-1996;
MAX PLANCK GEBIETSSCHAFT (DE)
COMMENT Other publication DE 4445562 960404.
FEATURES
source 1. 4137
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 1127 a 999 c 1001 g 1010 t
ORIGIN
Query Match 80.0%; Score 16; DB 5; Length 4137;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGNAGYATTCGNCARGA 20
|||:||||:||||:|
Db 2823 GGTGATGATTCGCGCAGA 2842

RESULT 8
LOCUS CEN97661 339 bp mRNA PLN 26-MAY-1998
DEFINITION Chlamydomonas eugametos phosphatidylinositol 3-kinase mRNA, partial
cds.
ACCESSION U97661
NID 92109286
VERSION U97661.1 GI:2109286
KEYWORDS
SOURCE Chlamydomonas eugametos.
ORGANISM Chlamydomonas eugametos
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE 1 (bases 1 to 339)
Molendijk, A.J. and Irvine, R.F.
TITLE Inositolide signalling in Chlamydomonas: characterization of a
phosphatidylinositol 3-kinase gene
JOURNAL Plant Mol. Biol. 37 (1), 53-66 (1998)
MEDLINE 98281574
REFERENCE 2 (bases 1 to 339)
Molendijk, A.J. and Irvine, R.F.
TITLE Direct Submission
JOURNAL Submitted (15-APR-1997) Pharmacology, University of Cambridge,
Tennis Court Road, Cambridge, United Kingdom
FEATURES
source 1. 339
/organism="Chlamydomonas eugametos"
/db_xref="taxon:3053"
/note="VP534 homolog"

CDS
CDS
1. >339
/note="VP534 homolog"
/product="phosphatidylinositol 3-kinase"
/protein_id="AAC50016.1"
/db_xref="PID:92109287"
/db_xref="GI:2109287"
/translation="GDDLRDQVLYQLFALMDRLRLRDVNLDRITPPYHLATSPDGL
VQFIPSVAAVAQVGGILSYLRFHFYEDHENATFHVQAOVLDTFVNSCAGYCVTY
LLGVGDRHLN"

BASE COUNT 68 a 102 c 94 g 75 t
ORIGIN
Query Match 80.0%; Score 16; DB 8; Length 339;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGNAGYATTCGNCARGA 20
|||:||||:||||:|
Db 1 GGGATGATTCGCGCAGA 20

RESULT 9
CRU97662 2200 bp mRNA PLN 26-MAY-1998
LOCUS Chlamydomonas reinhardtii phosphatidylinositol 3-kinase mRNA,
partial cds.
DEFINITION
ACCESSION U97662
NID G2109288
VERSION U97662.1 GI:2109288
KEYWORDS
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE 1 (bases 1 to 2200)
AUTHORS Molendijk,A.J. and Irvine,R.F.
TITLE Inositolide signalling in Chlamydomonas: characterization of a phosphatidylinositol 3-kinase gene
JOURNAL Plant Mol. Biol. 37 (1), 53-66 (1998)
MEDLINE 98281574
REFERENCES 2 (bases 1 to 2200)
AUTHORS Molendijk,A.J. and Irvine,R.F.
TITLE Direct Submission
JOURNAL Submitted (15-APR-1997) Pharmacology, University of Cambridge, Tennis Court Road, Cambridge, United Kingdom
FEATURES
source
1..2200
/organism="Chlamydomonas reinhardtii"
/strain="CW-15"
/db_xref="taxon:3055"
<1..2199
/note="VPS34 homolog"
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="AAC50017.1"
/db_xref="PID:92109289"
/db_xref="GI:2109289"
/translation="ESGAGAGSGPRAGGAIITLVNPEVGRDPAEIKAKKLRGICORCY
LDRDKPSEERRRQATILKIPNKPPIAEERALLMRASLTQDTRALTEFLQCDVM
SDATEAKQAALMRQWAPISADALELSPFSEVVRQAVQALQAVETELAYTL
QVQALREADDTGSRGAFVQRAAEDPALAVQHWLFCGLDPGGLRARAHLTL
LDLTLAGGSPRSGILLADITRQVDVYKQKAVIRELSSKIKARATEBLREMSV
TGAGGELLALVPLIDPHITVLAGIMPRECSVKSANPLRISPIEAPPMWKAAPP
PQNRHTNPNRPHRPHGLHAPAGASHSSGSSGSSGAVHSPGPGCGCGGG
WPPDGPALPLPTKRGILPAPGAPDAGAGGAGGAGGAGGAGGAGGAGGAGGAGG
ASRGAGGAGAPSGPAGPVPVLEGRCCIIRKGBDLQDHFVQLQIMLDMRMLREN
LDLRMTPKVLPVTFSDGLVEFVSPVSLAVLAERKITRFLATQADAGGPGFKRAE
AVETFRSCAGCYWMTYILGVDRHLDMMLTHGGRFLHIDFGYLGADTKPPPPMK
LCKEMTAMGGGSGEYKQPMYACENIIRKADLTLSFHLMAGASTAIRADNPE
KAILKLDKFRLLDDEDAVEMOGLINESATLDPULVETTHRAQYWR"
BASE COUNT 347 a 775 c 772 g 306 t
ORIGIN
Query Match 80.0%; Score 16; DB 8; Length 2200;
Best Local Similarity 65.0%; Pred. NO. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

RESULT 11
HSD1EM 5397 bp mRNA PRI 07-OCT-1996
LOCUS H.sapiens mRNA for phosphatidylinositol 3 kinase gamma.
DEFINITION
ACCESSION X83368
NID 91507821
VERSION X83368.1 GI:1507821
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 5397)
AUTHORS Stoyanov,B., Volinia,S., Hanck,T., Rubio,I., Loubtchenkov,M.,
Malek,D., Stoyanova,S., Vanhaesebroeck,B., Dhondt,R., Nuernberg,B.,
Gierschik,P., Seedorf,K., Hsuan,J.C., Waterfield,M.D. and
Eukaryota; Fungi; Ascomycota; Archiascomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
1 (bases 1 to 2634)
Takegawa,K., Denald,D.B. and Emr,S.D.
Schizosaccharomyces pombe Vps34p, a phosphatidylinositol-specific
PI 3-kinase essential for normal cell growth and vacuole morphology
J. Cell. Sci. 108 (Pt 12), 3745-3756 (1995)
JOURNAL 96360252
MEDLINE
REFERENCE 2 (bases 1 to 2634)
AUTHORS Takegawa,K., Denald,D.B. and Emr,S.D.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-1995) Scott Emr, Div. of Cellular and Molecular
Medicine, University of California at San Diego School of Medicine,
9500 Gilman Drive, San Diego, CA 92093-0668, USA
FEATURES
source
1..2634
/organism="Schizosaccharomyces pombe"
/db_xref="taxon:4896"
57..2462
/gene="Spvps34"
57..2462
/gene="Spvps34"
/note="Spvps34"
/codon_start=1
/product="Spvps34p"
/protein_id="AAC49133.1"
/db_xref="PID:929997"
/db_xref="GI:929997"
/translation="MDRLVPSYCPSSKVTARFLVFKPFIEYODSQEPCICTIOFSN
ESGSLMOKCTPSKIPKSLPTELKSTHEMLDQGVTSLSNAFVSAKPSRN
DEVEYEFVGGCTYRLFDENNLRLQGLKIPLOKSKETKTSPTSLKQVKEIRKLG
LLIQLOLGDVPSVMWDIDISGKIKDRSRHMSLVTPILYLFQKSPFVPEGRSY
KPSERNRYYSFDELNDLSDPAELKRLVRSQNGPLDLDKSKIRKESILSY
PSEBELSELEDLIMKFRFYLRNKKMRTFKLVVWTDSEVQALSLDSMTEDID
DDALELISPSVHPKVRVAVASRLKTSNEELLYLQVQALRYMDISSDEROPS
PLAFVYNRAISSSTINDUAKWIVYVEIEDEPVSUKLSYVHFLFKELKSVBRLLR
ETLQAQKFEKLIRISKVSQSEFGTLKRIEYKLVLEDRKTHLDLFDLPLDLS
VNIKEGILPDCTVYKSYMOPRLKLFKQDSSKYEPIFRNDDQLDQVLQIOLIMK
LTKERKIDLHKPRIATAGTLPKAVQFVSKTATIAEHGSLVLAILENNDDGL
NSAYNGZDPVADMDVRSQAGCYITVLLGVGDHLDMLITKNGHFHDFGYILGR
DKPFSAMLSKEMVSGMGVNSPFQOGRKSYTYTETALRSNNLINFSLMVA
NIPDKIKDKERVYKVERKLCQMSSEDAIKYFQILINDYSALFPQIIDRMENLAQY
MRS"
BASE COUNT 817 a 477 c 463 g 877 t
ORIGIN
Query Match 80.0%; Score 16; DB 8; Length 2634;
Best Local Similarity 65.0%; Pred. NO. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Wetker, R.
Cloning and characterization of a G protein-activated human
phosphoinositide-3 kinase
Science 269 (5224), 690-693 (1995)
JOURNAL MEDLINE 95350661
REFERENCE 2 (bases 1 to 5397)
AUTHORS Waterfield, M.D.
TITLE Direct Submission
SUBMITTED (08-DEC-1994) M.D. Waterfield, Ludwig Inst. for Cancer
Research, Courtauld Building, 91 Riding House Street, London, W1P
8BT, UK
REMARK 3 (bases 1 to 5397)
AUTHORS Waterfield, M.D.
TITLE Direct Submission
SUBMITTED (23-AUG-1996) M.D. Waterfield, Ludwig Inst. for Cancer
Research, Courtauld Building, 91 Riding House Street, London, W1P
8BT, UK
COMMENT On Aug 26, 1996 this sequence version replaced gi:940513.
X83368 is homologous to M93352.
FEATURES
source Location/Qualifiers
1. 5397
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="U937"
324..3629
/note="phosphatidylinositol 3 kinase gamma, p110 gamma,
activated by G protein alpha and betagamma subunits"
/codon_start=1
/product="idem"
/protein_id="CAA58284.1"
/db_xref="PID:664520."
/db_xref="PID:91507822"
/db_xref="GI:1507842"
/db_xref="SWISS-PROT:P48736"
/translation="MELENYKOPVLRREDCRRRRMRKPRSAASLSMELPIEFVLP
TSORCKSPETALHVAHGNGVOMKAOYALSTVADYHRLGPHHLLIYOK
GOWEYIDKQYVQVTLDCRYKAKTHRSQTHIVORHPSESOQAFQQLNALLGYD
VIDSVHDEDELEFRRLVTRMAEVAASRDKLVAHPWVSKPLPETLAKIANNC
IFIVHIRSTQSITIVSPDTPGAILQSEFTVMAKSLMDIPESQSDQFVLRKICGR
DEYLVGEPKINQVNRHCLKNGEIHVVLDPDPALDEVRKEEMPLVANDCTGYG
HEQLTIGHDESEFTVSLMDCDRKRVIRIGIDIPVLRNDLVEYVAVANQOQV
LCORRSPKPEVEYLMNMYLSEIKIKDLPFGALLNLIQYGRKAPASLSKASAPS
SSKGRVLLVYVNLIDHRLIRGEVYLMNMOISGGEQGSFNDKLSALNP
KENSMSITLNDNCHPLAPKHPQTPDEGVRKEMNOLKOLELAIADPLNPL
TADEKELMHFYESLKHPRKAYPKLESYVQOGEIVAKTYQLARREYMDSALDVG
LTMQLDCNFSDENRALVAVKLESDDDVLAHLYQVAVKFEFHDLSALRLK
RLRNKRIGHLEFLRLSEIAQSRHYQGRFVILAYLRGCTAMLDHETQOQVYEM
LOKVLDIRLSAEKYDSSOYISOLKOKLELONSOLPESFRVYDPQLKGAIAIE
KQVWASKKPKMLEKCAPALSNETGIIFFKHGDDLRQMLIOLIRINESHWET
ESLDLCLEPYGISTGDKIGMIEIVKDATITIKQSTVNGTNGARFDEVLNHSLKES
PTEEFKQAAVERFVSCAGYCVATVLAIGIDHNHNNIMTTEGNIHIDFGHLLANIK
SLGINKERVPVLPDLFLVMTGSKGKPHQKFOODICVAVYALRHHLLILFL
SMMLTGMPLTSEKEDIEYIRDALTVGKNEEDAKKAYFLDIEVCBKGWTVQFNMFLH
LVLGKQGEKHSNA"

BASE COUNT 1534 a 1224 c 1207 g 1432 t
ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 5397;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNAGAYATTCGNCARGA 20
|||:||||:|||||
Db 2823 GGTGATGATCTGGCCACAGA 2842

RESULT 12
LOCUS HSPH13K 3424 bp mRNA
DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
ACCESSION 229990

NID 9472990
VERSION 229990.1 GI:472990
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1069)
AUTHORS Volinia, S., Hiles, I., Ormondroyd, E., Nizetic, D., Antonacci, R.,
Molecular cloning, cDNA sequence, and chromosomal localization of
the human phosphatidylinositol 3-kinase p110 alpha (PIK3A) gene
Genomics 24 (3), 472-477 (1994)
JOURNAL MEDLINE 95229146
REFERENCE 2 (bases 1 to 3424)
AUTHORS Volinia, S.
TITLE Direct Submission
SUBMITTED (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig
Institute for Cancer, Research, 91 Riding House Street, London, W1P
8BT, UK
FEATURES
source Location/Qualifiers
1. 3424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="K562"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CAA82333.1"
/db_xref="PID:9472991"
/db_xref="GI:472991"
/db_xref="SWISS-PROT:P42336"
/translation="MPAPSSGELMGITLMPRLIVLECLLPNGMIVLECLREATLVY
IKHLEFKAERYPLHQLODESYIFVSTOEAAREDEFDRCLDLRLQPLKVI
EPVGRREKILNREIGFAIGMPVEEDVKNPEYODFERNTLNCKEAVDRDNPSPH
SRAMVYAPPHSEPLPKHYNLDRGOLLVIVVSPNNDOKQYTLKRNHCVPE
QVIAEAIKTRSHLSSEDLKLYLEIQTGRIIKVGCDEYLEKTPLSQYKTRSC
IMLGMPMLKMKAKESLSQLPMDCFMPSYSRIATPYMNETSTRKSLWVNRAL
RIKILCAVYVNLNRIDKIVRGFIYHGSPICDNNVOTVPCSNPKNMNLVYDIY
IPDPLRAARLILSISYSGRGAKEHCPLAWGVINLPDYDTLVSGMALNLMPVPH
GLEDLNLIGTGSVNPKEPTCLEEFPMFSSVVKFPMDSVIEEHANNSVREGFSY
SHAGLSVNDLNRENDKOLAISTRDLSITDQKFLMSHHRYCVTPEIILP
KLISVKNNSDENVDAQYKVMKPIRPEQAMLELDCNIPDPYRGAFAVCKEYIL
DDKSQYILQVLYKTEQYLDNLVRLKALFNRIQGFHFWHLSEKHNFTVSQ
RFEGVLESCACGNYLKLNRQVYAEKILNLIDILKQERKDETQKQKFLVEQNR
RPDMDALOGILSPNPAHQGNRLKECRIMSSAKRPLMNWNPDIIMSELEONNE
IIFKNGDLRQMDLQIIRILMENIMONOGDLMPLPGCLSIDGCVGLIIVNSHT
IMOIQCKGLGALQFNSHLMQKKNKEIYDAIDLPTRCAGACVATFTLIGIG
DRHNSNTVKKDGGQFLHDEGHFLDHKKKKRGRVRRVRYLQDFLVISKGAQECT
KTRERERQENQYKAYLAIRQHANLFINLFSNMLGSGPELOSFDDIAYIRKTLALDK
TEQALVEFMQNDADHGGWTTKMDWTFHTRIKQHALN"

BASE COUNT 1134 a 618 c 709 g 963 t
ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 3424;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNAGAYATTCGNCARGA 20
|||:||||:|||||
Db 2422 GGGAGATTTAGGCCACAGA 2441

RESULT 13
LOCUS HSP1TR1 2990 bp mRNA
DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
ACCESSION 246973
NID 9987947
VERSION 246973.1 GI:987947
KEYWORDS phosphatidylinositol 3-kinase.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
1 (bases 1 to 887)
AUTHORS Volinia, S., Dhand, R., Vanhaesebroeck, B., MacDougall, L.R., Stein, R.,
Zvelebil, M.J., Dhand, R., Panarecu, C. and Waterfield, M.D.
TITLE A human phosphatidylinositol 3-kinase complex related to the yeast
Vps34p-Vps13p protein sorting system
JOURNAL EMBO J. 14 (14), 3339-3348 (1995)
MEDLINE 95354652
REFERENCE 2 (bases 1 to 2990)
AUTHORS Volinia, S.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1994) Stefano Volinia PhD, Receptor Studies,
Ludwig Institute For Cancer, Research, 91 Riding House Street,
London, W1P 8BT, UK
FEATURES
source Location/Qualifiers
1..2990
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="p1R-1"
/cell_line="TF-1, KG1a"
/clone_1lb="Lambda zap rfl, KG1a"
1..47
48..2711
/product="phosphatidylinositol 3-kinase"
/protein_id="CA87094.1"
/db_xref="PID:987948"
/db_xref="GI:987948"
/db_xref="SPTREMBL:Q15134"
/translation="MGAEKFKHYISCDIDINVQLKIGSLGKREOKSYNAVLEDPML
KESGILQETCSLDLYTCQVFAEGKPSALPRTSYKAFSTEKMNNEMLKLPKPPDLP
NOVALITIMVYGVKAVPGGTVSLFEGYKSGROGMHLKVPVNEADSGPTVLP
GRSSLTSEDOMSLAKITRAHROGHVAVDMIDLPREIKEMINSESVKSNHMLRS
GGRVYKCDKDEKGYIYKEDGESSPILSFELVYPPDQSEMLEVSKHNMLRS
LRSGPDHDKLPSPRDLKNTVSPKSPPTERODLWERTITLINDOKLITIL
TSYIWLPGAKALALIKWNPMDVEDSLISSHYTNTVAVARLROADDL
LMTLSQVLAKEYENFDKNGDEPTKDSQSGVNSGNSAIEISSQITISPL
PSVSPSPASKTEVEPDGELNEDICTFLISRAKNSSTANLYLVYIVCEPDQIQ
ROKTEHMYLVNVRRESQALLKGDQSVRVRSLAAQOTFVRLVILMKRVOESGR
KKKNEKIALIGNEKWNLSDELIPLEPOYKIRGIIDPETATFKSALMPQLEPK
TEBGKTPYTFKGGDDLRODQLIQLIISLMDLKRKENDLKIPIKVLATSKHGK
QFQSVPAVEVDLTBGSIONFERKIVASEGPGKISAEVADIVKSCAGCVITTYLG
VGDRLDNLVLTGTGKLFHIDFGYIIGRDPKPLPPMKLNKEVBNCGTOSQYQEF
RKQCYATFLHRLRYSNLILNLFSLMVDNPIDIALPDKTKVKYQDKFRIDLSDERAV
HYWQSLIDESVALFAAVYEQIKFAQYMKR"
BASE COUNT 2712 ..2990 566 c 678 g 797 t
ORIGIN 3'UTR
3949 a 566 c 678 g 797 t
Query Match 80.0%; Score 16; DB 10; Length 2990;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
1 (bases 1 to 3207)
AUTHORS Stidivant, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledger, L.M.,
Cliff, A. and Heimbrock, D.C.
TITLE Catalytic Activity of the p110-alpha Subunit of Human
Phosphoinositide 3'-Hydroxylase is Required for Signal
Transduction
JOURNAL Bioorg. Med. Chem. (1996) In press
REFERENCE 2 (bases 1 to 3207)
AUTHORS Stidivant, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledger, L.M.,
Cliff, A. and Heimbrock, D.C.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs,
Summerville Pike, West Point, PA 19486, USA
FEATURES
source Location/Qualifiers
1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
/function="phospholipid kinase"
1..3207
/product="phosphoinositide 3'-hydroxylase p110-alpha
subunit"
/protein_id="BAB39753.1"
/db_xref="PID:q1763626"
/db_xref="GI:1763626"
/translation="MPRRSGELMGHILMPRIIVECLLPNGMIVTECLREATLIT
IKHELFEARXPPLHQLDDESSYIVSVTOAEREFEDERLCLDLPOPKVI
EYVNGEERKILNREIGAFGIMPCVCEPDQVNDQDFRNTINVCENADLNSPH
SRMYVYPPNVSSPELPKHITKDKQIIVYIVYVSPNNDKQYITIKTHDCYPE
QVIAEIRKTRSMLSLSBQLKLVLEYGKYLIVCGDEFFLEKYSQYKIRSC
IMGKPNMLMAKESLYSQLPMDQFTMSYSRRISTAPYNNKERTSLWVINSAL
RKIKIATLVANNIRIDIKIYRTGTYHSGEPLCDNVDNTPRCSNPRNENLVNPIY
IDPLPAAATGCTSGVKGKRAKHECHLAGNINILFYTDLVSKALNMLPVPH
GLEDILNPICGVNDSNPKNETPLETFEDPSSVYVPPQSVTEHANNVSGAGSY
SHAGLSNRLANDELRENDKEDKALISTDPISEITBOKDILNSRHVCYVITPELP
KLILSVKNSRDEVAQMTCLVDMPEIRKQAMELIDCVTPDPAVGRFVAVRLEKIT
DKRLSOYLILQVQLKYEQYLDNLVRFKRLKALNORIGHEFFVHMLKEMNKTVSQ
RGILLESYCRACGMYLKLNRQVEAMEKLINLIDILKQEKDEQVQMKFLVEQR
RDEMDLGGFSLPAPLQNLNRLRECRINSKRPPLMNNENDISELIFORNE
IIRKNDGLRQMLTIQIRIMENINNOGRLDRLMPCGLSIGCVGLIEVVRNFT
IMQIQCKGLKALQDNSTTHQWLKDKKGEITDPAIDLFRSCGCVVAVFIIIG
DRHNSINWVCKQYFLHIDFGHLDKRRKREIKRKEVFLVLDPLIVISGAQDCT
KTRFERPEQCKAYIALAIOHANKFENFNSMLSGKVELOSFDIAYIRITLADK
TDEALAEYMKQNDANHHGQWTYKMDWIFHTIKOHALN"
BASE COUNT 1043 a 586 c 670 g 908 t
ORIGIN
Query Match 80.0%; Score 16; DB 10; Length 3207;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

AUTHORS Hu, P., Mondino, A., Skolnik, E. Y. and Schlesinger, J.
TITLE Cloning of a novel, ubiquitously expressed human
phosphatidylinositol 3-kinase and identification of its binding
site on p85
JOURNAL Mol. Cell. Biol. 13 (12), 7677-7688 (1993)
MEDLINE 94067128
REMARK Genbank staff at the National Library of Medicine created this
entry (NCBI gi140879) from the original journal article.
This sequence comes from Fig. 1.

FEATURES

source

1. 3213

/organism="Homo sapiens"

/db_xref="taxon:9606"

gene

/gene="phosphatidylinositol 3-kinase p110 beta isoform, PI

CDS

1. 3213

/gene="phosphatidylinositol 3-kinase p110 beta isoform, PI

3-kinase p110 beta"

/note="110 kda catalytic subunit; This sequence comes from

fig. 1; PI 3-kinase p110 beta"

/product="phosphatidylinositol 3-kinase p110 beta isoform"

/protein_id="AAB29081.1"

/db_xref="pid:9455760"

/db_xref="gi:455760"

/translation="MCFSEIMPAMADLIDIVANDSQTASDGIPIVDFLLPGIYITOL
EVPREATISYIKOMLMKQVHNTPMNLMDIDSYMFACVNOYAVVELEDETRLCQV
RPEFLVLRKVRSCDPKELSKIGVLIGKGLHEDSLKDEPVNERRRKRFSEKI
LSIVGLSMWMDYMKQYPPPEHPSIPENTEDKLYGSKLIYAVFENCQDFSPVSPNM
NPDKYNELAIORLTIHGKEDVSPDYVLOVSGHVEYFGDHPILIOFYIRNCVMNR
ALPHILVECCCKIKMKYQEMATEAIAINRNSNPLPLPKKTIISHVWENNPRQ
IYVYGNKLNTEETVYKVRAGLPHGTTELCTIYSSEYSGKNDHTNPELFDINIC
DLPRARLCFAVYAVLDKVKTKKSTYQIRAKGVHVPVAVNTMVEDEKQ
QLRIGDILHSSSPDELEKLNPMGTQNTPTENATALHVKRPENKQPYYP
DKIIEKAAEIASDSANYSRGCKFLPYLKEIILDRPLSOLCEENMDLITLRQCR
EIEPOSILPLLSIKWKLEDAVQALLOIIMPKLPPREALELIDENYPDQYVREYAV
GCLRMSDEELSOYLLOLVOLKYEPPLDCAISRLEIRALGNRRIGQPLFMHLSBY
HIPAVSDFGVILEYCRGSYGHMKVYLSKQYDALNKLTLNSLILNVAKLNRAGKE
AMHTLCKOSAYREALSDLSPLNCPVILSELYVENCKYMSMKRPLMLVYNNKVFGE
SVGLIFKNGDDLRQDMITLQMLRLMDLMEKAGLDRLPYGCLATGDRSGLIEVYST
SETIADIOLNSNNVAAAFAFNKDALNMLKEVNSGDDLDRAIEEFTLSCAGCVASY
LGIGDRHSDNIVKTKGOLFHDIFGHILGNFRSKGIRKREYFILTIDPFIHVIOGK
TGNTEKGFGRFCCEDAYLILRRHGNLFTILPALMTAGLPELTISVKDIOYLKDSIAL
CKSEEARLQKQKQKDEALRESHTTKVMMAHITVAKDYRS"

BASE COUNT 979 a 612 c 704 g 918 t
ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 3213;

Best Local Similarity 65.0%; Pred. No. 1e+02; Mismatches 0; Gaps 0;

Matches 13; Conservative 5; Indels 0; Gaps 0;

DB 2419 GGTGATGATTACGACAGCA 2438

Search completed: September 13, 1999, 15:55:59
Job time: 4587 sec

FT /note="PI3- kinase p110"

PN MO9321328-A.

LU 28-OCT-1993.

PF 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDM-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI ParkerPJ, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

DR P-PSDB: R43341.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PS activity, useful for controlling cell proliferation

PS Claim 7; Fig 9: 146pp: English

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC prophylaxis or therapy. Platelet or neutrophil activity or blood

CC glucose levels can be controlled using the kinase.

CC See also Q51156, Q59012-23 and Q57522-3.

SQ Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match 80.0%; Score 16; DB 1; Length 3207;

Best Local Similarity 65.0%; Pred. No. 7.6;

Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNGAYGAYTTCGNCARGA 20

DB 2410 GGGAGACTTACGCGACAGA 2429

RESULT 3

O57522

ID 057522 standard; cDNA; 381 BP.

AC 057522.

DT 12-APR-1994 (first entry)

DE Human P1TR-C cDNA.

KW Phosphoinositide kinase; PI: p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KN ss.

PN MO9321328-A.

PD 28-OCT-1993.

PF 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDM-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI ParkerPJ, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

DR P-PSDB: R46552.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PS activity, useful for controlling cell proliferation

PS Disclosure: Fig 20; 146pp: English.

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC prophylaxis or therapy. Platelet or neutrophil activity or blood

CC glucose levels can be controlled using the kinase.

CC See also Q51155-6, Q59012-23 and Q57523.

SQ Sequence 381 BP; 116 A; 78 C; 102 T;

Query Match 80.0%; Score 16; DB 1; Length 381;

Best Local Similarity 65.0%; Pred. No. 5.7;

Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNGAYGAYTTCGNCARGA 20

DB 1 GGGAGACTTACGCGACAGA 20

RESULT 4

O57523

ID 057523 standard; cDNA; 393 BP.

AC 057523.

DT 12-APR-1994 (first entry)

DE Human P1TR-C cDNA.

KW Phosphoinositide kinase; PI: p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KN ss.

PN MO9321328-A.

PD 28-OCT-1993.

PF 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDM-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI ParkerPJ, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

DR P-PSDB: R46553.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PS activity, useful for controlling cell proliferation

PS Disclosure: Fig 21; 146pp: English.

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC prophylaxis or therapy. Platelet or neutrophil activity or blood

CC glucose levels can be controlled using the kinase.

CC See also Q51155-6, Q59012-23 and Q57522.

SQ Sequence 393 BP; 87 A; 103 C; 124 G; 79 T;

Query Match 80.0%; Score 16; DB 1; Length 393;

Best Local Similarity 65.0%; Pred. No. 5.7;

Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNGAYGAYTTCGNCARGA 20

DB 1 GGGAGACTTACGCGACAGA 20

RESULT 5

O59022

ID 059022 standard; DNA; 20 BP.

AC 059022.

DT 12-APR-1994 (first entry)

DE Degenerate primer for p110 cDNA.

KW Phosphoinositide kinase; PI: p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KN ss; amplification.

OS Synthetic.

PN W09321328-A.
 PD 28-OCT-1993.
 PR 13-APR-1993: G00761.
 PR 13-APR-1992: GB-008135.
 PA (LUDWIG) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otau M, Panayotou G,
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 2: Page 53; 146pp; English.
 CC Two novel cDNAs related to p110 have been cloned. Degenerate
 CC primers were designed to conserved sequences between human p110 and the
 CC related yeast gene VP534. These were used in RT-PCR using mRNA from the
 CC human cell lines MOLT4 and U937. Two novel cDNAs, P1TR-C and P1TR-F,
 CC related to p110, were isolated.
 CC See also 051155-6, 059012-23 and 057522-3.
 SO Sequence 20 BP; 4 A; 2 C; 6 G; 1 T;

Query Match 80.0%; Score 16; DB 1; Length 20;
 Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGNAGAYATTCGNCARGA 20
 DB 1 GGNAGAYATTCGNCARGA 20

RESULT 6
 057012
 ID 057012 standard: cDNA to mRNA; 3498 BP.
 AC 057012;
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase 110 kD catalytic subunit cDNA.
 KM 110 kD catalytic subunit: phosphatidylinositol 3-kinase;
 KM transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KM Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KM blood vessel plaques; ss.
 OS Bos taurus.
 FH Key
 FT cds 1. 3307
 FT /tag- a
 FT /product- p110

PN W09403609-A.
 PD 17-FEB-1994.
 PR 05-AUG-1993: G01651.
 PR 05-AUG-1992: GB-016654.
 PA (TMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PW, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 DR P-PSDB; R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure; Fig 1; 71pp; English.
 CC This sequence encodes the 110 kD catalytic subunit of the phosphatidylyl
 CC inositol (Ptdins) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nmt promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the ptdins
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nmt promoter functions and the ptdins 3-kinase
 CC catalytic subunit is induced. Ptdins activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SO Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 80.0%; Score 16; DB 1; Length 3498;

Best Local Similarity 65.0%; Pred. No. 7.7;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNAGAYATTCGNCARGA 20
 DB 2410 GCGGATGATTCGCGCAGA 2429

RESULT 7
 758545
 ID 758545 standard: cDNA; 4134 BP.
 AC 758545;
 DT 25-MAR-1997 (first entry)
 DE Human phosphatidylinositol 3-kinase PI3K-gamma cDNA.
 KM Phosphatidylinositol 3-kinase; gamma; PI3K; immunogen; immunocassay;
 KM cell proliferation; receptor-mediated signal transduction;
 KM histamine secretion; nerve cell differentiation; glucose transport;
 KM modulation; regulation; Alzheimer's disease; lipolysis; ds.
 OS Homo sapiens.
 FH Key
 FT cds 423. 3572
 FT /tag- a
 FT /product- PI3K-gamma

PN DE4445562-C1.
 PD 04-APR-1996.
 PR 20-DEC-1994: 445562.
 PR 13-OCT-1994: DE-436696.
 PA (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.
 PI Hanck T, Stoyanov B, Wetzel R;
 DR WPI: 96-172545/18.
 DR P-PSDB; W11576.
 PT New phosphatidylinositol 3-kinase protein - useful as immunogen and
 PT for determ. of kinase activity
 PS Claim 5; Page -: 10pp; German.
 CC A 402 bp cDNA fragment was amplified from a human bone marrow
 CC library using PCR primers corresponding to amino acid sequences
 CC KNGDDR and HIDEG. The amplified fragment was used to probe a human
 CC cDNA library and several overlapping clones were isolated.
 CC The largest clone had the present sequence and coded for a protein
 CC of 1049 residues. The protein is a novel phosphatidylinositol 3-
 CC kinase (PI3K) that differs in its regulatory mechanism from the
 CC known PI3K-alpha and -beta enzymes. The new enzyme has been
 CC designated PI3K-gamma and can be used as an immunogen. The enzyme,
 CC antibodies against it or nucleic acid encoding it can be used for
 CC modulating cell proliferation, receptor-mediated signal transduction,
 CC histamine secretion, nerve cell differentiation, glucose transport
 CC and anti-lipolytic activity or for treating Alzheimer's disease.
 CC N.B. Although the claimed sequences are referred to by SEQ. ID.
 CC numbers, a sequence listing did not appear in the original printed
 CC patent application.
 SO Sequence 4134 BP; 1127 A; 999 C; 998 G; 1010 T;

Query Match 80.0%; Score 16; DB 1; Length 4134;
 Best Local Similarity 65.0%; Pred. No. 7.8;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNAGAYATTCGNCARGA 20.
 DB 2620 GCGGATGATTCGCGCAGA 2639

RESULT 8
 758546
 ID 758546 standard: cDNA; 4137 BP.
 AC 758546;
 DT 25-MAR-1997 (first entry)
 DE Human phosphatidylinositol 3-kinase PI3K-gamma cDNA.
 KM Phosphatidylinositol 3-kinase; gamma; PI3K; immunogen; immunocassay;
 KM cell proliferation; receptor-mediated signal transduction;
 KM histamine secretion; nerve cell differentiation; glucose transport;
 KM modulation; regulation; Alzheimer's disease; lipolysis; ds.
 OS Homo sapiens.

FH Key Location/Qualifiers
 FT cds 423..3575
 FT /*tag= a
 FT /product= PI3K-gamma
 PN DE4445562-C1.
 PD 04-APR-1996.
 PR 20-DEC-1994: 445562.
 PR 13-OCT-1994: DE-436696.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PI Hancock T, Stoyanov B, Wetzker R;
 DR WPI: 96-172545/18.
 DR P-PSDB: M11577.
 PT New phosphatidylinositol 3-kinase protein - useful as immunogen and for degerm. of kinase activity
 PS Claim 5, Page -: 10pp; German.
 CC A 402 bp cDNA fragment was amplified from a human bone marrow library using PCR primers corresponding to amino acid sequences CC KNGDDR and HIDEF. The amplified fragment was used to probe a human CC 0937 cell cDNA library and several overlapping clones were isolated. CC The largest clone coded for a protein of 1049 residues. The protein CC is a novel phosphatidylinositol 3-kinase (PI3K) that differs in its CC regulatory mechanism from the known PI3K-alpha and -beta enzymes. CC The new enzyme has been designated PI3K-gamma. Another clone, coding CC for a PI3K-gamma of 1050 residues and having the present CC sequence, was then isolated. The enzyme can be used as an immunogen. CC The enzyme, antibodies against it or nucleic acid encoding it can be CC used for modulating cell proliferation, receptor-mediated signal CC transmission, histamine secretion, nerve cell differentiation, CC glucose transport and anti-lipolytic activity or for treating CC Alzheimer's disease.
 CC N.B. Although the claimed sequences are referred to by SEQ.ID, CC Numbers, a sequence listing did not appear in the original printed CC patent application.
 SQ Sequence 4137 BP; 1127 A; 999 C; 1001 G; 1010 T;

Query Match 80.0%; Score 16; DB 1; Length 4137;
 Best Local Similarity 65.0%; Pred. No. 7.8;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNGAYGAYTRCGNCARGA 20
 DB 2823 GGTGATGATCTGCCCANAGA 2842

RESULT 9
 ID V04634
 AC V04634:
 DT 17-AUG-1998 (first entry)
 DE Porcine phosphoinositide 3OH-kinase p120 subunit cDNA.
 KW G-beta-gamma regulated phosphatidylinositol-3' kinase; pig;
 KW phosphoinositide 3OH-kinase; PI3K; signal transduction;
 KW phosphatidylinositol (3,4,5)-triphosphate; G-protein; receptor;
 KW transgenic animal; knockout animal; inflammation; arthritis;
 KW septic shock; adult respiratory distress syndrome; pneumonia;
 KW asthma; allergy; reperfusion injury; atherosclerosis; cancer;
 KW Alzheimer's disease; cancer; antisense; ribozyme; diagnosis;
 KW therapy; drug screening; ss.
 OS Sus scrofa.
 FH Key Location/Qualifiers
 FT CDS 89..3397
 FT /*tag= a
 PN M09749818-A2.
 PD 31-DEC-1997.
 PR 26-JUN-1997: U11219.
 PR 27-JUN-1996: US-672211.
 PA (ONXX-) ONXX PHARM.
 PI Braselmann S, Hawkins PT, Stephens L;
 DR WPI: 98-077181/07.
 DR P-PSDB: W23948.
 PT DNA encoding G-beta-gamma regulated phosphatidylinositol-3' kinase, p101 and p120 subunits - useful for diagnosis, drug screening,

PT clinical trial monitoring and treatment of inflammatory disorders
 PS Claim 54; Fig 3A-C; 151pp; English.
 CC This cDNA sequence codes for the p120 catalytic subunit (see CC W23948) of pig G-protein regulated phosphatidylinositol-3' kinase, CC a heterodimeric enzyme which produces the intracellular messenger CC phosphatidylinositol (3,4,5)-triphosphate in response to activation CC of trimeric G protein-linked receptors. This novel protein, which CC also contains a regulatory subunit, p101 (see V04633), is found in CC cells of hematopoietic origin and is involved in immune system CC responses which cause inflammation. p120 cDNA was obtained from a CC pig neutrophil cDNA library using a degenerate oligonucleotide CC probe (see T99713) based on an isolated p120 tryptic peptide. The CC p120 cDNA clone in pCMV3mycP120 is deposited as ATCC 97637. The CC invention encompasses pig and human p101 and p120 nucleotides, CC host cell expression systems, polypeptides and peptides, antibodies CC W23946-49), fusion proteins, polypeptides and peptides (see CC to these proteins, and transgenic animals and knockout animals. CC Compounds which are useful for treating inflammatory response CC disorders can be identified by screening assays using a G protein CC activated PI3K, or a cultured host cell that expresses the p101 CC gene. Antagonists of G protein stimulated PI3K (acting through the CC p101 subunit, especially by disrupting the interaction between the CC p101 and p120 subunits) can be used to treat arthritis, septic CC shock, adult respiratory distress syndrome (ARDS), pneumonia, CC asthma, allergies, reperfusion injury, atherosclerosis, cancer and CC Alzheimer's disease. The nucleic acids and their products can also CC be used for diagnosis, drug screening and clinical trial monitoring CC of inflammatory diseases.
 SQ Sequence 3808 BP; 1010 A; 965 C; 953 G; 880 T;

Query Match 80.0%; Score 16; DB 1; Length 3808;
 Best Local Similarity 65.0%; Pred. No. 7.7;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNGAYGAYTRCGNCARGA 20
 DB 2591 GGTGAGATCTGCCCANAGA 2610

RESULT 10
 ID V74100
 AC V74100:
 DT 09-MAR-1999 (first entry)
 DE Porcine G-protein regulated PI3K p120 adapter subunit DNA.
 KW Phosphoinositide 3-hydroxylase; PI3K; trimeric G protein; porcine;
 KW adapter subunit; regulatory subunit; p101; p120; catalytic subunit;
 KW detection; diagnosis; activation disorder; hematopoietic system;
 KW treatment; immune disorder; inflammation; arthritis; septic shock;
 KW adult respiratory distress syndrome; pneumonia; asthma; allergy;
 KW reperfusion injury; atherosclerosis; Alzheimer's disease; cancer; ss.
 OS Sus scrofa.
 PN US5856132-A.
 PD 05-JAN-1999.
 PR 15-AUG-1997: 916917.
 PR 15-AUG-1997: US-916917.
 PR 27-JUN-1996: US-672211.
 PA (ONXX-) ONXX PHARM.
 PI Braselmann S, Hawkins PT, Stephens L;
 DR WPI: 99-105107/09.
 DR P-PSDB: W90083.
 DR Nucleic acid encoding regulatory (p101) and catalytic (p120) PT subunits of a heterodimeric phosphatidylinositol-3' kinase - useful PT in treatment and diagnosis of immune system disorders, e.g. PT arthritis, cancer and Alzheimer's disease
 PS Example IX; Fig 3A-C; 75pp; English.
 CC This sequence encodes a novel catalytic subunit, p120, from porcine CC phosphoinositide 3-hydroxylase (PI3K) which is regulated by beta-gamma CC subunits of trimeric G-protein. Nucleic acid coding for p101 or p120, or CC their fragments, are used as probes and primers for identifying p101 or CC p120 gene mutations, allelic variations or regulatory defects, CC particularly for the diagnosis of activation disorders (or

CC susceptibility) in cells of the haematopoietic system. The related
 CC proteins, antibodies, agonists and antagonists can be used similarly. The
 CC p101 and p120 proteins, peptides or fusion proteins are used to treat or
 CC screen for potential agents for treating immune disorders, particularly
 CC inflammation, e.g. arthritis, septic shock, adult respiratory distress
 CC syndrome, pneumonia, asthma, allergies, reperfusion injury,
 CC atherosclerosis, Alzheimer's disease and cancer.
 SQ Sequence 3808 BP; 1010 A; 965 C; 953 G; 880 T;

Query Match 80.0%; Score 16; DB 1; Length 3808;
 Best Local Similarity 65.0%; Pred. No. 7.7;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNAGYGYTRCGNCARGA 20
 |||:||||:|||||
 DB 2591 GGTGACGATCTCGCCCAAGA 2610

RESULT 11
 V74104
 ID V74104 standard; DNA: 5162 BP.

AC V74104;
 DT 09-MAR-1999 (first entry)
 DE Human G-protein regulated PI3K p120 adapter subunit DNA.
 KW Phosphoinositide 3-hydroxylase; PI3K; trimeric G protein; human;
 KW adapter subunit; regulatory subunit; p101; p120; catalytic subunit;
 KW detection; diagnosis; activation disorder; haematopoietic system;
 KW treatment; immune disorder; inflammation; arthritis; septic shock;
 KW adult respiratory distress syndrome; pneumonia; asthma; allergy;
 KW reperfusion injury; atherosclerosis; Alzheimer's disease; cancer; ss.
 OS Homo sapiens.
 PN US856132-A.
 PD 05-JAN-1999.
 PF 15-AUG-1997; 916917.
 PR 15-AUG-1997; US-916917.
 PR 27-JUN-1996; US-672211.
 PA (ONTX-) ONTX PHARM.
 PI Braseleiman S, Hawkins PT, Stephens L;
 DR WPI: 99-105107/09.
 DR P-PSDB: W90089.
 PT Nucleic acid encoding regulatory (p101) and catalytic (p120)
 PT subunits of a heterodimeric phosphatidylinositol-3' kinase - useful
 PT in treatment and diagnosis of immune system disorders, e.g.
 PT arthritis, cancer and Alzheimer's disease
 PS Example X: Fig 12A-C; 75pp; English.
 CC This sequence encodes a novel catalytic subunit, p120, from human
 CC phosphoinositide 3-hydroxylase (PI3K) which is regulated by beta-gamma
 CC subunits of trimeric G-protein. Nucleic acid coding for p101 or p120, or
 CC their fragments, are used as probes and primers for identifying p101 or
 CC p120 gene mutations, allelic variations or regulatory defects,
 CC particularly for the diagnosis of activation disorders (or
 CC susceptibility) in cells of the haematopoietic system. The related
 CC proteins, antibodies, agonists and antagonists can be used similarly. The
 CC p101 and p120 proteins, peptides or fusion proteins are used to treat or
 CC screen for potential agents for treating immune disorders, particularly
 CC inflammation, e.g. arthritis, septic shock, adult respiratory distress
 CC syndrome, pneumonia, asthma, allergies, reperfusion injury,
 CC atherosclerosis, Alzheimer's disease and cancer.
 SQ Sequence 5162 BP; 1471 A; 1173 C; 1149 G; 1367 T;

Query Match 80.0%; Score 16; DB 1; Length 5162;
 Best Local Similarity 65.0%; Pred. No. 8.1;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNAGYGYTRCGNCARGA 20
 |||:||||:|||||
 DB 2586 GGTGATGATCTCGCCCAAGA 2605

RESULT 12
 060936

ID 060936 standard; DNA: 423 BP.

AC 060936;
 DT 16-MAR-1994 (first entry)
 DE Human brain Expressed Sequence Tag EST01019.
 KW Gene transcription product; genetic markers; tagging; in vivo;
 KW transcription; mapping; locations; chromosomes; chromosomal; ss.
 OS Homo sapiens.
 PN W09316178-A.
 PD 19-AUG-1993.
 PF 12-FEB-1993; U01294.
 PR 12-FEB-1992; US-837195.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Adams MD, Moreno RF, Venter CJ;
 DR WPI: 93-272882/34.
 PT Enriched oligonucleotides and corresp. sequences - used as
 PT markers for human genes transcribed in-vivo, facilitate tagging
 PT of most human genes
 PS Example 4: Page 413; 500pp; English.
 CC The Expressed Sequence Tag was isolated from a human brain cDNA
 CC library as part of a large set of ESTs which can be used as markers
 CC for human genes transcribed in vivo. They can be used to facilitate
 CC tagging of most human genes, for mapping locations of expressed genes
 CC on chromosomes, for individual or forensic identification, for mapping
 CC locations of disease-associated genes, for identification of tissue
 CC type, and for prep. of antisense sequences, probes and constructs.
 CC EST01019 has an "excellent" coding probability as evaluated using the
 CC coding-region prediction program CRM. See also 059041-061440.
 SQ Sequence 423 BP; 87 A; 122 C; 124 G; 89 T;

Query Match 74.0%; Score 14.8; DB 1; Length 423;
 Best Local Similarity 65.0%; Pred. No. 27;
 Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGNAGYGYTRCGNCARGA 20
 |||:||||:|||||
 DB 272 GAGACGACTTCGCCCAAGA 291

RESULT 13

ID T96605/c
 AC T96605; standard; DNA: 1612 BP.
 DT 19-FEB-1998 (first entry)
 DE C. glutamicum atrop gene.
 KW Corynebacterium glutamicum; atrop gene; transporter; aromatic amino acid;
 KW microbial amino acid production; secretion rate; tryptophan; tyrosine;
 KW phenylalanine; feed additive; aspartame; ss.
 OS Corynebacterium glutamicum ATCC 13032.
 FH Key Location/Qualifiers
 FT CDS 80..1471
 FT /*tag= a
 PM DE19523279-A1.
 PD 09-JAN-1997.
 PF 27-JUN-1995; 023279.
 PR 27-JUN-1995; DE-023279.
 PA (KERU) FORSCHUNGSZENTRUM JUELICH GMBH.
 PI Eggeling L, Kraemer R, Wehrmann A;
 DR WPI: 97-066157/07.
 DR P-PSDB: W36452.
 PT Microbial prodn. of (aromatic) amino acids in cell transformed with
 PT gene for transport protein - provides increased secretion of product
 PT into culture medium
 PS Claim 8; Page 5; 7pp; German.
 CC The present sequence contains the Corynebacterium glutamicum ATCC 13032
 CC atrop gene which encodes a transporter protein for aromatic amino acids.
 CC This gene is useful in the microbial production of amino acids. The
 CC expression of the gene increases the secretion rate of the amino acid
 CC and hence the cell excretes a greater amount of the amino acid into the
 CC medium. In particular microbial production of e.g. tryptophan (for
 CC pharmaceuticals or feed additives), tyrosine (for pharmaceuticals or
 CC intermediates) or phenylalanine (an intermediate for aspartame), is
 CC preferred.

SO Sequence 1612 BP; 280 A; 450 C; 421 G; 461 T;

Query Match 74.0%; Score 14.8; DB 1; Length 1612;

Best Local Similarity 65.0%; Pred. No. 32;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGAGYGYTTCGNCARGA 20

DB 235 GGGAGTGTACGCCAGCA 216

RESULT 14

VO1877

VO1877 standard: DNA; 3602 BP.

VO1877;

20-APR-1998 (first entry)

DE Human NP1K gene.

KM Human; foetal brain cDNA library; GDP dissociation stimulating protein;

KM Brain specific nucleosome assembly protein; diagnosis; therapy;

KW skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NP1K;

KM nel-related protein type 1; nel-related type 2; hereditary disease;

KW cancer; ss.

OS Homo sapiens.

FT Key

Location/Qualifiers

429..2882

CDS

/*tag- a

/product- "NP1K"

EP-796913-AZ.

PD 24-SEP-1997.

PF 19-MAR-1997; 104842.

PR 05-MAR-1997; JP-069163.

PA 19-MAR-1996; JP-063410.

PA (SARA) OTSUKA PHARM CO LTD.

PI Fujiwara T, Horie M, Watanabe T;

DR WPI: 97-459830/43.

P-PSDB; W37498.

PT Novel human genes, e.g. brain-specific nucleosome assembly protein -

PT useful for diagnosis or therapy of hereditary disease and cancer

PS Claim 15; Page 78-83; 123pp; English.

CC The present sequence encodes a NP1K isolated from a human foetal brain

CC cDNA library. The nucleotide or amino acid sequences are useful for

CC in-vitro diagnosis of hereditary diseases and cancer and for preparation

CC of pharmaceuticals.

SO Sequence 3602 BP; 849 A; 1006 C; 985 G; 762 T;

Query Match 74.0%; Score 14.8; DB 1; Length 3602;

Best Local Similarity 65.0%; Pred. No. 35;

Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGAGYGYTTCGNCARGA 20

DB 2124 GGGAGTGTACGCCAGCA 2143

RESULT 15

VO1878

VO1878 standard: cDNA; 2487 BP.

VO1878;

20-APR-1998 (first entry)

DE Human NP1K gene.

KM Human; foetal brain cDNA library; GDP dissociation stimulating protein;

KM Brain specific nucleosome assembly protein; diagnosis; therapy;

KW skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NP1K;

KM nel-related protein type 1; nel-related type 2; hereditary disease;

KW cancer; ss.

OS Homo sapiens.

FT Key

Location/Qualifiers

1..2487

CDS

/*tag- a

/product- "NP1K"

/note- "no stop codon given"

PN EP-796913-AZ.

PD 24-SEP-1997.

PF 19-MAR-1997; 104842.

PR 05-MAR-1997; JP-069163.

PA 19-MAR-1996; JP-063410.

PA (SARA) OTSUKA PHARM CO LTD.

PI Fujiwara T, Horie M, Watanabe T;

DR WPI: 97-459830/43.

P-PSDB; W37499.

PT Novel human genes, e.g. brain-specific nucleosome assembly protein -

PT useful for diagnosis or therapy of hereditary disease and cancer

PS Claim 17; Page 87-88; 123pp; English.

CC The present sequence encodes a NP1K isolated from a human foetal brain

CC cDNA library. The nucleotide or amino acid sequences are useful for

CC in-vitro diagnosis of hereditary diseases and cancer and for preparation

CC of pharmaceuticals.

SO Sequence 2487 BP; 604 A; 682 C; 664 G; 537 T;

Query Match 74.0%; Score 14.8; DB 1; Length 2487;

Best Local Similarity 65.0%; Pred. No. 34;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGAGYGYTTCGNCARGA 20

DB 1732 GGGAGTGTACGCCAGCA 1751

Search completed: September 13, 1999, 15:59:37
Job time: 4804 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:57 ; Search time 1694.61 Seconds
(without alignments)
23.280 Million cell updates/sec

US-09-325-095-27

Title: perfect score: 20
Sequence: 1 GGNAGYATTCGNCARCA 20

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	16	80.0	324	20	T29663	T29663 EST89481 Hu
2	16	80.0	424	22	H14711	H14711 YM24A02.r1
3	16	80.0	375	28	C12533	C12533 C12533 Yu1
4	16	80.0	667	30	AA206574	AA206574 zq51902.t
5	16	80.0	597	30	AA210248	AA210248 m084c09.t
6	16	80.0	437	35	AA542224	AA542224 VK28A04.r
7	16	80.0	357	35	AA571816	AA571816 v191f09.t
8	16	80.0	574	37	AB009107	AB009107 AB009107
9	16	80.0	413	39	AA661836	AA661836 ak39A08.S
10	16	80.0	497	47	AI531818	AI531818 SP03087.S
11	14.8	74.0	423	20	M78871	M78871 EST01019 Su
12	14.8	74.0	271	20	Z42324	Z42324 HSCOFF081.n
13	14.8	74.0	223	23	D57845	D57845 HUM429H07B
14	14.8	74.0	396	24	D61665	D61665 HUM428B12B
15	14.8	74.0	413	30	AA231198	AA231198 mw38f12.r
16	14.8	74.0	475	33	AA437822	AA437822 ve34g12.r
17	14.8	74.0	727	41	AI1001270	AI1001270 MEST5-D1.
18	14.8	74.0	750	42	AI158022	AI158022 pd26A08.r
19	14.8	74.0	443	46	AI426863	AI426863 mm97e11.x
20	14.8	74.0	538	51	AI714581	AI714581 605059G11
21	14.4	72.0	384	21	R17530	R17530 yg03d11.r1
22	14.4	72.0	593	24	N24991	N24991 yk16e11.r1
23	14.4	72.0	419	26	W95377	W95377 ze01g12.r1
24	14.4	72.0	451	27	AA001396	AA001396 ze64a05.r
25	14.4	72.0	280	27	AA017578	AA017578 ze37f04.r
26	14.4	72.0	231	31	AA329679	AA329679 EST33318
27	14.4	72.0	441	33	AA452699	AA452699 zK39D09.r
28	14.4	72.0	334	34	AA485157	AA485157 aa40A05.r
29	14.4	72.0	599	38	AA801073	AA801073 EST190570
30	14.4	72.0	302	39	AA842951	AA842951 T4689 MVA
31	14.4	72.0	206	40	AT000268	AT000268 AT000268
32	14.4	72.0	95	44	AI298931	AI298931 gm59p03.x
33	14	70.0	507	37	AA698421	AA698421 H104464.5
34	14	70.0	784	41	AI069387	AI069387 mg2e006d
35	14	70.0	349	49	AU055795	AU055795 AU055795
36	14	70.0	167	49	AV017180	AV017180 AV017180
37	13.8	69.0	103	20	D12085	D12085 HUM0516A08
38	13.8	69.0	764	30	AA264673	AA264673 LP08378.5
39	13.8	69.0	512	31	AA310586	AA310586 EST181407
40	13.8	69.0	264	33	AA433315	AA433315 TE0F0148
41	13.8	69.0	360	36	C62862	C62862 C62862 Yu1
42	13.8	69.0	360	36	C64513	C64513 C64513 Yu1
43	13.8	69.0	375	36	C70607	C70607 C70607 Yu1
44	13.8	69.0	362	42	AI145747	AI145747 UT-R-BF0-
45	13.6	68.0	424	25	W13743	W13743 mb32c10.r1

ALIGNMENTS

RESULT 1
T29663 LOCUS T29663 324 bp mRNA EST
DEFINITION EST89481 Human Small Intestine Homo sapiens cDNA 5' end similar to
phosphatidylinositol 3-kinase p110, beta isoform (HT3411), mRNA
ACCESSION T29663
NTD 9611761

VERSION T29663.1 GI:611761
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 324)
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
 Bult,C.J., Lee,N., Kirkness,E.F., Weissbrock,K.G., Gocayne,J.D.,
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.W.,
 Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
 Fitzgerald,L.M., Fitzhugh,M.M., Fritchman,J.L., Geoghegan,N.S.M.,
 Glodet,A., Gnehm,C.L., Hanna,M.C., Hedblom,J.E., Hinkley,P.S.,
 Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.T., Marmaros,S.M.,
 Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,
 Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
 Seudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
 Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A.,
 Coleman,T.A., Collins,E.-J., Dinko,D., Feng,P., Fierle,A.,
 Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,D.M.,
 Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Li,H., Li,H.,
 Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wang,J., Xu,C.,
 Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
 Hesselme,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
 Initial Assessment of Human Gene Diversity and Expression Patterns
 Based Upon 83 Million Basepairs of cDNA Sequence
 Nature 377, 3-174 (1995)
 96026280

TITLE
 JOURNAL
 MEDLINE
 COMMENT

CONTACT: Venter, JC
 The Institute for Genomic Research
 932 Clopper Rd, Gaithersburg, MD 20878
 Tel: 3018690956
 Fax: 3018699423
 Email: tdbinfo@db.tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please contact the TIGR Database
 (tdbinfo@db.tigr.org)
 Seq primer: M13 Reverse.

FEATURES
 Location/Qualifiers
 source
 1..324
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):107054"
 /db_xref="taxon:9606"
 /clone_lib="Human Small Intestine"
 /note="Organ: small intestine"
 BASE COUNT 88 a 58 c 80 g 96 t 2 others

Query Match 80.0%; Score 16; DB 20; Length 324;
 Best Local Similarity 65.0%; Pred. No. 43;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNAGYAGYTRCGNCARGA 20
 |||:||||:|||||
 Db 70 GGTGATGATTACGACAGCA 89

RESULT 2
 LOCUS H14711 424 bp mRNA EST 27-JUN-1995
 DEFINITION YM24602.F1 Soares Infant brain INIB Homo sapiens cDNA clone
 IMAGE:48772 5' similar to SP:P100_BOVIN P22871 PHOSPHATIDYLINOSITOL
 3'-KINASE CATALYTIC SUBUNIT ;, mRNA sequence.
 H14711
 NID 9879531
 VERSION H14711.1 GI:879531
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 424)
 Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
 Wilson,R.
 The WashU-Merck EST Project
 Unpublished (1995)
 On Sep 21, 1992 this sequence version replaced gi:276288.

TITLE
 JOURNAL
 COMMENT

CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1696
 High quality sequence strops: 312
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1696 Std Error: 0.00
 Seq primer: M13Rpl
 High quality sequence stop: 312.

FEATURES
 Location/Qualifiers
 source
 1..424
 /organism="Homo sapiens"
 /db_xref="GDB:421313"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:48772"
 /clone_lib="Soares Infant brain INIB"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: Lactid BA; Site: 1: Not
 I; Site 2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer [5',
 AACTGAGACAAATTCGCGCGCCAGCAATTTTCTTTTCTTTT 3'];
 (double-stranded cDNA was ligated to Hind III adaptors
 (pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the Lactid BA vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 107 a 85 c 104 g 124 t 4 others

Query Match 80.0%; Score 16; DB 22; Length 424;
 Best Local Similarity 65.0%; Pred. No. 47;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNAGYAGYTRCGNCARGA 20
 |||:||||:|||||
 Db 198 GGTGATGATTACGACAGCA 217

RESULT 3
 LOCUS C12533 375 bp mRNA EST 28-DEC-1998
 DEFINITION C12533 Yui1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
 clone yk152b12 5', mRNA sequence.
 C12533
 NID 91560086
 VERSION C12533.1 GI:1560086
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabdilitia;
 Rhabdilitia; Rhabdilitidae; Rhabdilitidae; Pelodieridae; Caenorhabditis.
 1 (bases 1 to 375)
 Kohara,Y., Mochizashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
 Sano,M., Miyata,A. and Nishigaki,A.
 Expression map of the C.elegans genome

JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1288981.

Contact: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.

FEATURES

1. 375
 Location/Qualifiers
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(el489)"
 /db_xref="taxon:6239"
 /clone_lib="Yuji Kohara unpublished cDNA"
 /note="dev_stage=varied, sex=Hermaphrodite male,
 tissue_type=whole animal"

BASE COUNT 114 a 61 c 79 g 115 t 6 others
 ORIGIN

Query Match 80.0%; Score 16; DB 28; Length 375;
 Best Local Similarity 65.0%; Pred. No. 45;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 GGNAGAGATTCGNCARCA 20
 |||:::|||||
 Db 11 GGAGATGATTTACGACAGA 30

RESULT 4
 LOCUS AA206574 667 bp mRNA EST 12-MAR-1998
 DEFINITION Zg51902.r1 Stratiogene neuroepithelium (#937231) Homo sapiens CDNA
 clone IMAGE:645170 5' similar to SW:P11B.HUMAN P42338
 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, BETA ISOFORM ;
 mRNA sequence.

ACCESSION AA206574
 NID 91801954
 VERSION AA206574.1 GI:1801954
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 667)
 Hillier,D., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
 Martin,J., Moore,B., Scheinberg,K., Steptoe,M., Tan,F.,
 Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WashU-NCI human EST Project
 * Unpublished (1997)
 On Apr 14, 1993 this sequence version replaced gi:692719.

CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 2550 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 297.

FEATURES

1. 667
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="GDB:5215421"
 /db_xref="taxon:9606"
 /map="688H07; 2"

/clone="IMAGE:645170"
 /clone_lib="Stratiogene neuroepithelium (#937231)"
 /dev_stage="Ntera-2/RA neuroepithelial cells"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SR-; Site:1: EcoRI; Site:2:
 XhoI; Cloned unidirectionally. Primer: Oligo-dT, NT2
 cells (Ntera-2/cl.D1) induced with Retinoic acid for 24
 hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5'
 adaptor sequence: 5' GAATTCGACGACAG 3' -3' adaptor
 sequence: 5' CTCGACTTTTCTTTTCTTTT 3'"
 BASE COUNT 176 a 139 c 162 g 179 t 11 others
 ORIGIN

Query Match 80.0%; Score 16; DB 30; Length 667;
 Best Local Similarity 65.0%; Pred. No. 54;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 GGNAGATTCGNCARCA 20
 |||:::|||||
 Db 213 GGTGATGATTTACGACAGA 232

RESULT 5
 LOCUS AA210248 597 bp mRNA EST 29-JAN-1997
 DEFINITION m084c09.r1 Beddingington mouse embryonic region Mus musculus CDNA
 clone IMAGE:560272 5' similar to TR:G987948 G987948
 PHOSPHATIDYLINOSITOL 3-KINASE. ; mRNA sequence.

ACCESSION AA210248
 NID 91807546
 VERSION AA210248.1 GI:1807546
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 597)
 Marra,M., Hillier,D., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Scheinberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HMT Mouse EST Project
 * Unpublished (1996)
 On Sep 12, 1996 this sequence version replaced gi:1394729.

CONTACT: Marra M/Mouse EST Project
 WashU-HMT Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:341064
 Possible reversed clone: similarity on wrong strand
 Seq primer: -28M13 rev1 EF from Amersham
 High quality sequence stop: 338.

FEATURES

1. 597
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL6 x DBA"
 /db_xref="taxon:10090"
 /clone="IMAGE:560272"
 /clone_lib="Beddingington mouse embryonic region"
 /tissue_type="embryo"
 /sex="pooled"
 /dev_stage="7.5dpc"
 /lab_host="DH12S"
 /note="Organ: whole embryo; Vector: PCMV-SPORT; Site:1:
 SalI; Site:2: NotI; Cloned unidirectionally. Primer:

Oligo dt. Gastrulating embryos were collected at 7.5dpc from C57BL/6 x DBA matings, excluding embryos that had developed head folds and all extraembryonic tissues. Average insert size: 1.3 kb (range: 0.5 - 3.0 kb). Referenced in Development 121, 2479-2489 (1995)*

BASE COUNT 169 a 134 c 156 g 138 t
ORIGIN

Query Match 80.0%; Score 16; DB 30; Length 597;
Best Local Similarity 65.0%; Pred. No. 52;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNGAGATTCGCGCARGA 20
DB 535 GGAGAGACTTCGTCACAGA 554

RESULT 6
AA542224

LOCUS 437 bp mRNA EST 30-JUL-1997
DEFINITION VK8a04.r1 Soares mouse mammary gland NBMG Mus musculus cDNA clone IMAGE:947886 5' similar to SM:PIB_HUMAN P42338 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, BETA ISOFORM ;

ACCESSION AA542224
NID 92288658
VERSION AA542224.1 GI:2286658
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. I (bases 1 to 437)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Thelsting, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HM1 Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397965.

Contact: Marra M/Mouse EST Project
WashU-HM1 Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:544742
Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 414.

FEATURES

SOURCE

1. 437
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:947886"
/clone_lib="Soares mouse mammary gland NBMG"
/tissue_type="mammary gland"
/sex="male"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7n3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5' TGTTACCAATCTGAATGCGAGCGCGCGAATGATGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into.

the Not I and Eco RI sites of the modified pT7n3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Patricia Bonaldo.*

BASE COUNT 103 a 111 c 123 g 99 t 1 others
ORIGIN

Query Match 80.0%; Score 16; DB 35; Length 437;
Best Local Similarity 65.0%; Pred. No. 47;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNGAGATTCGCGCARGA 20
DB 183 GGTGAGATTCGCGCAGAGA 202

RESULT 7
AA571816

LOCUS 357 bp mRNA EST 27-AUG-1997
DEFINITION v191f09.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA clone IMAGE:988073 5' similar to TR:G987948 G987948 PHOSPHATIDYLINOSITOL 3-KINASE ; mRNA sequence.

ACCESSION AA571816
NID 92346745
VERSION AA571816.1 GI:2346745
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. I (bases 1 to 357)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Thelsting, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HM1 Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Apr 14, 1993 this sequence version replaced gi:315350.

Contact: Marra M/Mouse EST Project
WashU-HM1 Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:560353
Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 147.

FEATURES

SOURCE

1. 357
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:988073"
/clone_lib="Stratagene mouse diaphragm (#937303)"
/tissue_type="diaphragm"
/sex="adult"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally from mRNA prepared from diaphragm muscle. Primer: Oligo dt. Average insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3'." CTCGAGTTTTTTTTTTTTTTTTTTT 3'." 74 c 86 g 90 t

REFERENCE Muscomorpha: Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 497)
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
Borstein, P., Lewis, S., and Rubin, G. M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948480.

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
hit genomic sequence AC005639
Plate: 30 row: H column: 3
High quality sequence stop: 476.

FEATURES
source location/Qualifiers

1..497
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture POT2"
/lab_host="DH5-alpha"
/note="Vector: POT2; Site-1: EcoRI; Site-2: XhoI; Sized
fractionated cDNAs were directly ligated into POT2.
Plasmid cDNA library."
BASE COUNT 125 a 143 c 125 g 104 t
ORIGIN

Query Match 80.0%; Score 16; DB 47; Length 497;
Best Local Similarity 65.0%; Pred. No. 49;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNAGAYTTCGNCARGA 20
|||:|:|:|:|:|:|:|:|:|
Db 371 GGATGATCTGCGCAGCA 390

RESULT 11 423 bp mRNA EST 26-MAY-1992
LOCUS M78871
DEFINITION EST01019 Subtracted Hippocampus Strata gene (cat. #936205) Homo
sapiens cDNA clone HHCPA26, mRNA sequence.
ACCESSION M78871
NID 9273185
VERSION M78871.1 GI:273185
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 423)
AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,
Ullrich, T.R., Nagle, J.W., Fields, C. and Venter, J.C.
TITLE Sequence identification of 2,375 human brain genes
JOURNAL Nature 355, 633-634 (1992)
MEDLINE 92168112
COMMENT Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
Seq primer: M13 forward.
Location/Qualifiers
1..423
/organism="Homo sapiens"
/db_xref="ATCC (thost):78337"

/db_xref="GDB:D0S1850E"
/db_xref="taxon:9606"
/clone_lib="HHCPA26"
/clone_lib="Subtracted Hippocampus, Strata gene (cat.
#936205)."
/note="Vector: lambdaZAP-II; The hippocampus library (#4
above) was substracted with a fibroblast cell line cDNA
library (Strata gene cat. #936209; W138 lung fibroblast
cell line; oligo-dT + random primed cDNA synthesis;
lambdaZAP-II vector, 1.0kb average insert size.) by the
method of Sive & St. John (Nucl. Acids Res. 16:10937,
1988)."
BASE COUNT 87 a 122 c 124 g 89 t 1 others
ORIGIN

Query Match 74.0%; Score 14.8; DB 20; Length 423;
Best Local Similarity 65.0%; Pred. No. 2,2e+02;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGNAGAYTTCGNCARGA 20
|||:|:|:|:|:|:|:|:|:|
Db 272 GGAGACGACTTCGCGCAGCA 291

RESULT 12 271 bp mRNA EST 09-NOV-1994
LOCUS Z42324
DEFINITION HSCOFF081 normalized infant brain cDNA Homo sapiens cDNA clone
c-0ff08, mRNA sequence.
ACCESSION Z42324
NID 9565737
VERSION Z42324.1 GI:565737
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 271)
AUTHORS Auffray, C., Behar, G., Bols, F., Bouchier, C., da Silva, C.,
Devignes, M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B.,
Lorenzo, F., Mitchell, H., Marlage-Samson, R., Pletu, G., Pouliot, Y.,
Sebastien, Kabakchis, C. and Tessier, A.
TITLE IMAGB: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT On Sep 21, 1992 this sequence version replaced gi:279243.

Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_lib1: y1c-universal.
Seq primer: (-21)M13 universal.
Location/Qualifiers

1..271
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="c-0ff08"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lambda ZAP; Site-1: HindIII;
Site-2: NotI; sex=Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue-type-total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the

BASE COUNT 61 a 83 c 75 g 51 t 1 others
 ORIGIN lafmid BA vector. Clone library from B.Souares, Psychiatry
 Dept. Columbia University, USA. Normalization_method:
 Bento Soares, P.N.A.S in press"

Query Match 74.0%; Score 14.8; DB 20; Length 271;
 Best Local Similarity 65.0%; Pred. No. 1.9e+02;
 Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGNAGYATYTCGNCARGA 20
 |||:||||:|||||
 Db 249 GGGATGACCTTCGGCARGA 268

RESULT 13
 D57845 223 bp mRNA EST 28-AUG-1995
 LOCUS
 DEFINITION HM329H07B Clontech human aorta polyA+ mRNA (#6572) Homo sapiens
 CNA Clone GEN-329H07 5', mRNA sequence.

ACCESSION D57845
 NID 9964467
 VERSION D57845.1 GI:964467
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 223)
 Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,
 Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H.,
 Takaiuchi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y.,
 Makawa, H., Shin, S. and Nakamura, Y.
 Fujiwara et al. (1995)
 Unpublished (1995)

TITLE
 JOURNAL
 COMMENT On May 9, 1995 this sequence version replaced gi:802631.

CONTACT: Tsutomu Fujiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawuchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035
 Insert length: 1248 Std Error: 0.00
 High quality sequence stop: 349.

FEATURES
 source
 1. 223
 Location/Qualifiers

BASE COUNT 86 a 38 c 50 g 49 t
 ORIGIN /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="19"
 /clone="GEN-329H07"
 /clone_11b="Clontech human aorta polyA+ mRNA (#6572)"

Query Match 74.0%; Score 14.8; DB 23; Length 223;
 Best Local Similarity 65.0%; Pred. No. 1.8e+02;
 Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGNAGYATYTCGNCARGA 20
 |||:||||:|||||
 Db 143 GGGACGAGTACGCGACGA 162

RESULT 14
 D61665 396 bp mRNA EST 14-DEC-1995
 LOCUS
 DEFINITION HM428B12B Clontech human fecal brain polyA+ mRNA (#6535) Homo
 sapiens CDNA clone GEN-428B12 5', mRNA sequence.
 ACCESSION D61665
 NID 9971029

VERSION D61665.1 GI:971029
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 396)
 Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,
 Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H.,
 Takaiuchi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y.,
 Makawa, H., Shin, S. and Nakamura, Y.
 Fujiwara et al. (1995)
 Unpublished (1995)

TITLE
 JOURNAL
 COMMENT On Apr 14, 1993 this sequence version replaced gi:765805.

CONTACT: Tsutomu Fujiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawuchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035
 Insert length: 1028 Std Error: 0.00
 High quality sequence stop: 234.

FEATURES
 source
 1. 396
 Location/Qualifiers

BASE COUNT 90 a 101 c 107 g 98 t
 ORIGIN /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="17 cen-pter: 17q21"
 /clone="GEN-428B12"
 /clone_11b="Clontech human fetal brain polyA+ mRNA
 (#6535)"

Query Match 74.0%; Score 14.8; DB 24; Length 396;
 Best Local Similarity 65.0%; Pred. No. 2.1e+02;
 Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGNAGYATYTCGNCARGA 20
 |||:||||:|||||
 Db 108 GGGATGACCTTCGGCARGA 127

RESULT 15
 AA231198 413 bp mRNA EST 26-FEB-1997
 LOCUS
 DEFINITION mw38f12.T1 Soares mouse 3ME12.5 Mus musculus cDNA clone
 IMAGE:673007 5' similar to gp:k62534 HIGH MOBILITY GROUP PROTEIN
 HMG2 (HUMAN); gb:246757 M.musculus mRNA for high mobility group 2
 protein (MOUSE);, mRNA sequence.

ACCESSION AA231198
 NID 91853555
 VERSION AA231198.1 GI:1853555
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.

1 (bases 1 to 413)
 Maira, M., Hallier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The Mashu-HMI Mouse EST Project
 Unpublished (1996)

CONTACT: Maira M/Mouse EST Project
 Mashu-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800

Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:412711
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 23.

FEATURES

Source

1. 413
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6j"
 /db_xref="taxon:10090"
 /map="788H12:14924.3"
 /clone="IMAGE:673007"
 /clone_11b="Scars mouse 3NME12 5"
 /sex="unknown"
 /tissue_type="fetus"
 /dev_stage="12.5dpc total fetus"
 /lab_host="DH10B"
 /note="Organ: whole fetus; Vector: pRT3D-Pac (Pharmacia)
 with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 15' TGTACCAATCTGAAGTGGAGCGCGCCCTATTATTTTATTTT
 3', on total mouse RNA [provided by Minoru Ko, Wayne
 State Univ.]; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pRT3 vector.
 Library went through one round of normalization, and was
 constructed by Bento Soares and M. Felima Bonaldo."

BASE COUNT 108 a 98 c 118 g 89 t
 ORIGIN

Query Match 74.0%; Score 14.8; DB 30; Length 413;
 Best Local Similarity 65.0%; Pred. No. 2.1e+02;
 Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

07 1 GGNGAYGAYTTCGNCARCA 20
 92 GGTGAAGACTTGCCACAGA 111
 40

Search completed: September 13, 1999, 15:45:57
 Job time: 4045 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Sequence - nucleic search, using sw model

Run on: September 13, 1999, 15:04:48 ; Search time 329.34 Seconds

(without alignments)
560,084 Million cell updates/sec

Title: US-09-325-095-1214.SEO

I

Perfect score: 58
Sequence: 1 AARATGAGATGATGATTTTCA.....TTTCATATGATTTTGCCCA 58

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database:

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB	ID	Description
1	33	56.9	38	5	A37242	Sequence 11

2	31.8	54.8	3452	4	AF001076
3	31.8	54.8	3389	17	AF001075
4	30.2	52.1	3207	3	BOVPHOS3KN
5	30.2	52.1	3498	5	A37232
6	30	51.7	1696	5	A17975
7	28.6	49.3	3424	10	HSPH13K
8	28.6	49.3	3207	10	HSU79143
9	28.6	49.3	3207	12	MMU03279
10	26.8	46.2	1957	17	LEUG6PD
11	26.6	45.9	3868	10	HSP110DEL
12	26.6	45.9	3525	10	HSU57843
13	26.6	45.9	5220	11	HSU6453
14	26.4	45.5	3600	7	SPCC1672
15	25.8	44.5	35820	36	CEL721E12
16	25.8	44.5	35103	37	CEL2DNEINH
17	25.6	44.1	2028	5	E03267
18	25.6	44.1	1347	5	E03268
19	25.6	44.1	2028	5	E05210
20	25.6	44.1	1347	5	E05211
21	25.6	44.1	2028	5	I72356
22	25.6	44.1	1367	5	I72357
23	25.6	44.1	2028	9	HUHPH8R1
24	25.6	44.1	1367	9	HUHPH8R2
25	25.6	44.1	1167	11	AF037351
26	25.4	43.8	95930	9	HS332011
27	25.4	43.8	22720	36	CEC34B7
28	25.2	43.4	197082	11	AC006249
29	25.2	43.4	23960	36	CEC54C8
30	25	43.1	128330	9	HS11DP6
31	25	43.1	81984	10	HS130E17
32	25	43.1	3132	12	MMU08587
33	24.8	42.8	3213	10	S67334
34	24.8	42.8	124351	11	AC005185
35	24.8	42.8	178707	35	AC007435
36	24.8	42.8	24159	42	AC005677
37	24.6	42.4	68726	7	AB017063
38	24.6	42.4	84710	7	AB026643
39	24.6	42.4	187755	34	AC006718
40	24.6	42.4	31310	36	CELF0E5
41	24.6	42.4	3712	36	DMPHOS3K1
42	24.4	42.1	16450	1	MTV051
43	24.4	42.1	2037	5	I41352
44	24.4	42.1	119001	8	ATRC005896
45	24.4	42.1	179599	11	AC006305

ALIGNMENTS

RESULT	1	A37242	38 bp	DNA	PAT	05-MAR-1997
LOCUS	A37242	Sequence 11 from Patent WO9403609.				
DEFINITION	A37242					
ACCESSION	g2294353					
VERSION	A37242.1	GI:2294353				
KEYWORDS						
SOURCE	unidentified.					
ORGANISM	unclassified.					

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. 38
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 7 a 6 c 8 g 11 t 6 others
ORIGIN

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. 38
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 7 a 6 c 8 g 11 t 6 others
ORIGIN

Query Match 56.9%; Score 33; DB 5; Length 38;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GATGATGGCCARCTGTTTCATATGATGATTTGGCCA 58
 |||
 Db 1 GATGATGGCCARCTGTTTCATATGATGATTTGGCCA 35

RESULT 2
 AF001076 3452 bp mRNA VRL 08-JUL-1997
 LOCUS Gallus gallus phosphonositide 3-kinase catalytic subunit mRNA,
 DEFINITION complete cds.
 ACCESSION AF001076
 NID 92245505
 VERSION AF001076.1 GI:2245505
 KEYWORDS
 SOURCE Chicken.
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
 Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 3452)
 AUTHORS Chang, H.W., Aoki, M., Furman, D., Auger, K.R., Bellacosa, A.,
 Tschilis, P.N., Cantley, L.C., Roberts, T.M. and Vogt, P.K.
 TITLE Transformation of chicken cells by the gene encoding the catalytic
 subunit of PI 3-kinase
 JOURNAL Science 276 (5320), 1848-1850 (1997)
 MEDLINE 97334438
 REFERENCE 2 (bases 1 to 3452)
 AUTHORS Chang, H.W., Aoki, M. and Vogt, P.K.
 TITLE Direct Submission
 JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
 Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
 BCC-239, La Jolla, CA 92037, USA
 FEATURES
 source 1..3452
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /rname="G1:2245505"
 /rsize_type="brain"
 1..150
 151..3354
 /gene="c-p3k"
 /note="p3k proto-oncogene"
 151..3357
 /note="c-p3k"
 /codon_start=1
 /product="phosphonositide 3-kinase catalytic subunit"
 /protein_id="AAB62534.1"
 /db_xref="PID:92245506"
 /translation="MPPRPSGELMGHILMPRIIVLCELLPNGMIVTLECLREATLIT
 IKHELEKARKYPLQLODESSYIFSVTOAEEREFDETRLODLPFLKVI
 EGVGRREKILNRREIGALGMICEDFMDKREODERNNITNCKEADLDANAP
 SNAIVCPNVSSELPKRTINKDKGIIYIVYIVSPNDKOKYTKIKHDCVPE
 OVIAIRKTRISMLSSQLKCYLEVGKYLIVCGDELLERYTKIKHDCVPE
 IMLRPPIIMAKESLYQLDPLDTFMPYSRISTAPYNNKGTASLMTINSAL
 RLICATYVNNIDIKIYVYTGIGSEPLCDVNTQRPSCNPNMELSDYM
 IDLPRAALGSLISVKGKRAKEHCPLEAGNINMEYDTOLVSGKALNLMVPH
 GLEDLNPIGVSNDPKETPCLELEDFMFSNPKYRPMQTVIEHANMTISELQPNY
 SYAGSNRIARPNELRESKEDLRCTDPLSEITBOEKDLKSHRHKVTPETLP
 KILSVKMSRDEVAOMTCLVDPMPKPEQAMELLDQYDPDPAFAVRCLELYLT
 DKLQYLYQLQVLYEQYLDNOLVRLKALNORIGHEFVHLKEMNKTVSQ
 RGLLESEYCRAGMYIKHLSROVEAMEKILNLTDLKOEKDEQOKVMKFLVQMR
 RDEMDALOGFISPLNPAHQLNLEECRIMSASAKPLMNMENDISELLFQMR
 IIFKNGDRLQMLTLOIRIMENIMONGDILRMIPYGLSLSDGVGLIEVRSHT
 IMQIOCKGKALPNSHTLQMLNDRKNGEKIDALDLFTRSCGICVATFLLIGIG
 DRHNSINIKDQGLFHDGFLDKRKKFGYKRRVPEVLTODFLIVISGAEQCT
 KTRPEQEMCKAYLAIRHANTLINFISMLISGMEPELOSFDIAYIRKTLALDK
 TQEALEVFMRKMDAHHGGMWKTKMDWIHTIKQHLN"
 3358..3452

BASE COUNT 1073 a 672 c 783 g 924 t
 ORIGIN

Query Match 54.8%; Score 31.8; DB 4; Length 3452;
 Best Local Similarity 81.1%; Pred. No. 0.36;
 Matches 30; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 22 ACATGATGGCCARCTGTTTCATATGATGATTTGGCCA 58
 |||
 Db 2921 AAGATGATGGACACCTGTTTCATATGATGATTTGGCCA 2957

RESULT 3
 AF001075 3389 bp mRNA VRL 08-JUL-1997
 LOCUS Avian sarcoma virus 16 gag-v-phosphonositide 3-kinase catalytic
 DEFINITION subunit fusion protein (gag-v-p3k) and retroviral env mRNAs,
 partial cds.
 ACCESSION AF001075
 NID 92245502
 VERSION AF001075.1 GI:2245502
 KEYWORDS
 SOURCE Avian sarcoma virus 16.
 ORGANISM Avian sarcoma virus 16.
 REFERENCE 1 (bases 1 to 3389)
 AUTHORS Chang, H.W., Aoki, M., Furman, D., Auger, K.R., Bellacosa, A.,
 Tschilis, P.N., Cantley, L.C., Roberts, T.M. and Vogt, P.K.
 TITLE Transformation of chicken cells by the gene encoding the catalytic
 subunit of PI 3-kinase
 JOURNAL Science 276 (5320), 1848-1850 (1997)
 MEDLINE 97334438
 REFERENCE 2 (bases 1 to 3389)
 AUTHORS Chang, H.W., Aoki, M. and Vogt, P.K.
 TITLE Direct Submission
 JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
 Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
 BCC-239, La Jolla, CA 92037, USA
 FEATURES
 source 1..3389
 /organism="Avian sarcoma virus 16"
 /db_xref="taxon:60629"
 <1..3252
 /gene="gag-v-p3k"
 /note="fusion gene of gag and v-p3k"
 1..78
 /note="gag-v-p3k"
 /note="derived from gag gene"
 <1..3252
 /gene="gag-v-p3k"
 /codon_start=1
 /product="gag-v-phosphonositide 3-kinase catalytic
 subunit fusion protein"
 /protein_id="AAB62532.1"
 /db_xref="PID:92245503"
 /translation="QQLIRTAPESTLTPGELIIRYLVDRKASHMPRIIVLCELLPNG
 MIVTELEKREALLIKHELEKARKYPLQLODESSYIFSVTOAEEREFDETR
 RCLDRLPQPLKVEIPNGREKILNRREIGALGMICEDFMDKREODERNNITNCKEADLDANAP
 VCKEAVDRLNANPVSALVYCPNVSSELPKRTINKDKGIIYIVYIVSPNDKOKYTKIKHDCVPE
 KOKYTKIKHDCVPEQYVIAEIRKTRISMLSSQLKCYLEVGKYLIVCGDELLERYTKIKHDCVPE
 LEKTPSOKTRIRSCIMLRKPNMLMAKESLYQLDPLDTFMPYSRISTAPYNNKGTASLMTINSAL
 GEATKSLMTINSALRIRILICATYVNNIDIKIYVYTGIGSEPLCDVNTQRPSCNPNMELSDYM
 CSNPWNMLSDYMTIDLPRAALGSLISVKGKRAKEHCPLEAGNINMEYDTOLVSGKALNLMVPH
 IISGKALNLMVPHGLDPLNPIGVSNDPKETPCLELEDFMFSNPKYRPMQTVIEHANMTISELQPNY
 EHAQNTLIRELGFNYSVAGLSNRARDELRESKEDLRCTDPLSEITBOEKDLKSHRHKVTPETLP
 KSHRHKVTPETLPKILSVKMSRDEVAOMTCLVDPMPKPEQAMELLDQYDPDPAFAVRCLELYLT
 MYRAVRCLELYLTDLKQYLYQLQVLYEQYLDNOLVRLKALNORIGHEFVHLKEMNKTVSQ
 FVHLKEMNKTVSQRGLLESEYCRAGMYIKHLSROVEAMEKILNLTDLKOEKDEQOKVMKFLVQMR
 ETOKVMKFLVQMRDEMDALOGFISPLNPAHQLNLEECRIMSASAKPLMNMENDISELLFQMR
 ENPDIMSELQONNEIIFKNGDRLQMLTLOIRIMENIMONGDILRMIPYGLSLSDGVGLIEVRSHT
 GDVGLIEVRSHTIMQIOCKGKALPNSHTLQMLNDRKNGEKIDALDLFTRSCGICVATFLLIGIG

```

misc_feature
      79..84
      /gene="gaq-v-p3k"
      /note="sequence of unknown origin"
misc_feature
      85..3252
      /gene="gaq-v-p3k"
      /note="derived from c-p3k gene"
      3250..3370
      /codon_start=1
      /product="retroviral env"
      /protein_id="AA62533.1"
      /db_xref="PID:g2245504"
      /db_xref="GI:2245504"
      /translation="MARNS"
BASE COUNT      1071 a      655 c      751 g      912 t
ORIGIN
Query Match      54.8%; Score 31.8; DB 17; Length 3389;
Best Local Similarity 81.1%; Pred. No. 0.36;
Matches 30; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY      22 ACGATGATGCCARCTGTTTCATATGATTTGGCCA 58
      1||||||| :|||:|||||:|||||
Db      2816 AAGATGATGACAACTGTTTCATATTGACTTTGGCCA 2852
      |||||:|||||:|||||:|||||:|||||
RESULT 4
BOVPHOS3KN      3207 bp      mRNA      MAM      19-AUG-1992
LOCUS      Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete
DEFINITION
ACCESSION      M93252
NID      g163519
VERSION      M93252.1
KEYWORDS      GI:163519
SOURCE      phosphatidylinositol 3-kinase.
ORGANISM      Bos taurus
      Bos taurus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
      Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
      Bovinae; Bos.
      1 (bases 1 to 3207)
      Hiles, I.D., Otsu, M., Vollinia, S., Fry, M.J., Gout, I., Dhan, R.,
      Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F.,
      Hsuan, J.Y., Courtneidge, S.A., Parker, P.O. and Waterfield, M.D.
      Phosphatidylinositol 3-kinase: Structure and expression of the 110
      kd catalytic subunit
      Cell 70, 419-429 (1992)
FEATURES
      source
      1..3207
      /organism="Bos taurus"
      /db_xref="taxon:9913"
      /cell_line="SCBAR-1"
      1..3207
      /EC_number="2.7.1.67"
      /codon_start=1
      /product="Phosphatidylinositol 3-kinase"
      /protein_id="AAA30696.1"
      /db_xref="PID:g163520"
      /db_xref="GI:163520"
      /translation="MPRRSSSGELMGHTAMPRLIVLELNGMIVILECRLRLTIT
      IKHLEFBEARKPLHQLQDDSSITFVSITAEKEEFEDTRICDLRFQPLKVIY
      EPVNNREKILNREIGFALIGMPVCEFDMPKQPEVDPRNLINCKEAVDRLDINSPE
      SRAYVYPPNVESEPELKHIVKRLDKGQILVIVLVIVSPNNDKQYTLKINHCVSE
      OVLEAIKRTKRMULSSEQLKLCVLEQGVILKVGCGDPELEKYLSQYKIRSG
      IMGRPMILMAKRESLYSOLPMDCITPMSYSRSTSTAPMGNGESTKSLMVISNAR
      RIKLCATYVAVNRIDDKITVIRGITHGGEPDLDNNTQVPPCSNPRMWNLMNTDIL
      IPDLPRLARILGTSVKGKRGAAEEHCPPLAMGNINLEFDYIDTLVSGKALNMLPVP
      GLDELNDIGTGSNPNKEITPCLLELEDFWFSVYKIPMSYIEEHANNSVREGFSP

```

Query Match	52.1%	Score 30.2	DB 3	Length 3207
Best Local Similarity 78.4%		Pred. No. 1.3		
Matches 29; Conservative 5; Mismatches 3; Indels 0; Gaps 0;				
QY 22	ACGATGATGCCACATCTGTCACATWATTTGGCCA 58			
Db 2771	AAGATGATGCACACTGTTTCATATAGATTTTGGACA 2807			
RESULT 5				
A37232	A37232 3498 bp DNA PAT 05-MAR-1997			
LOCUS	Sequence 1 from Patent WO9403609.			
DEFINITION	A37232			
ACCESSION	9229435			
NID	A37232.1 GI:2294345			
VERSION	unidentified.			
KEYWORDS	unclassified.			
SOURCE	unclassified.			
ORGANISM	1 (bases 1 to 3498)			
REFERENCE	Parker, P. J., Goode, N. T., Nurse, P. M. and Waterfield, M. D.			
AUTHORS	EDUCAROTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE			
TITLE	OR PROTEIN KINASE AND ASSAYS USING THEM			
JOURNAL	Patent: WO 9403609-A 1 17-FEB-1994;			
COMMENT	IMP CANCER RES TECH (GB)			
FEATURES	Other publication JP 8503124T 960409.			
source	Location/Oualifiers			
	1. 3498			
	/organism="unidentified"			
	/db_xref="taxon:32644"			
	1. 3207			
CDS	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="Caa02284.1"			
	/db_xref="PID:e304808"			
	/db_xref="PID:g2294346"			
	/db_xref="GI:2294346"			
	/translation="MPPRPSGSELMG;HLMPPRLVCLLPNGMIYTLDELRENTLIT IKHELFKARKYPLPHQLDSESYIEFSVQEEEREFEDTRLCDRLRFLPKAYI EYGVNREKILNREIGFALIGMPVECDPMVADPEVDFRRNMLVNCXAVDLRLDSPE SRAYVVPWVESPELPHKILYNKLDGQILVAVLVSVNNKOKYTLKINDXSE OVAEAIARKTRSMLESEDLKLCVLEQKYLKVGCEYFLEKYLKSOYKIRSC IMGKRNMLMKRESYSQLPMDCTFMPYSRILSNTPYNGESTKSLMYNSL RIKLCATYVAVNIRIDDKLIVRTGITHGSEPLCDNNTVORPCSNRRNEMLYDIL IPDLPRARLCLSTCSYKGRKGAEEHCPLAWGINLFDYTLVSSKMLNLMPVH GLDELNLIGVTSNPKETPCLELEDNWSYVKPDMSVIEEHLVSSMLNMPVH SHAGLSNRLARNDLENDEKOLRAICTRDLSEITFEQKDFLMSHHYCVTIPELP KLILSVKNSDEVAOMYCLVKMDPPIKPRQAMELDCDNFDPVNRGEAVRCLEKYL DRLTSOYLIOUYLKYEOYLDNLIVLPELTKLTLNORIGHPFMHILKSMHNKYVSQ RFGILSEYCRACGMYLKLHNRQVEAMEKILNLTDLKQKDETCVQKFLVEONR RPFMDLQGLFSLPNPAHQNLNRECHIMSSAKRPLMLNMDNPIMSELEHONNE IIFENGDDLQDMQLTQILIRIMENIMONQGLDRLMLPYGCLSGDCVGLLEVNSHT IMOLOCGGLKALQLOPNSHTLHOKLWKNQGEIYDAIDLFTSCAGYCAVTEILIG DRHNSNIMVDDQGLFHDHGHFLDKHKKKFKGKREVPVYLQDPLIYVSKAOEOT KTRFEFEQEMCKTAYLAIKQHANLFINLSSMLGSGMPELOSFDDIATYIRKTLALDK TEQDALEYFKOMKDANHGMGTMMDFIHTIKOHALN"			
BASE COUNT	1028 a 581 c 680 g 918 t			
ORIGIN				
BASE COUNT	1147 a 626 c 722 g 1003 t			

/gene="p110delta"
 197. 3331
 /gene="p110delta"
 /codon_start=1
 /product="phosphoinositide 3-kinase"
 /protein_id="CAA71149.1"
 /db_xref="PID:c308053"
 /db_xref="PID:g2104840"
 /db_xref="GI:2104840"
 /db_xref="SWISS-PROT:000329"
 /translation="MPPGVDCPMEFVETKINQVLEEOLEDEORRLCVQFPLVRLRY
 IKOLLWRADQEPFLFHLISGPFAVFTICINQVLEEOLEDEORRLCVQFPLVRLRY
 ARGDRYKPLINSQISLLIGKLGHEPDSLDLPVNFPRAMCOEFCSEAARPOOLGE
 AWLYQSEPIOLEPESAOITMGSTLRILPNRALVYKFKGSESESTFOVSTDAVPLAMA
 CARAKATVPRIOLVPODEYDTQVNGRHHYLGSDVLCOPVYICSLTHSLPLHTM
 VHSSTILAMDEDSNPAPOVQYKPRAKPPILVPCPKSSVLSMSLEDPRIELTIGQSKM
 ADRMKLVVQAGLPHNGEMLCRTVSSSEVSVCSEPMQRLEFDINICDLPNARLCF
 ALVALEKAKARSTKRSKQACDPIAMANMLDYKDKQLEKRLYMWPSVDEKQ
 ELINPQTVASNPNTSSAALLICLPEVAHPVYALERILETGRSECEVYAEQEO
 LQREILERSGEELVEHERDILVWKLRLHEVQHEPELRLALLVTKNKRHEVYAOQMEY
 LQCSWELPYLALIELIDPSPDCHGSSPFIKSLRKTIDDELFOYLLOITVOVLKYESY
 LDEELIKFLIDRALNARKTIGHFLRSMHPHVSVALRPGTILAEACRGRTHHMKY
 MKEGELSKTKALANDYKLSQKTPKQYRELMHLCKRQATYLEALSHLOSPEDPSYL
 LAVEVQOCTFMDSKMKPLIMTISNEPAGSGSVGLIFKNGDYLDDMTLQMLQMD
 VLMKQGLDRMPDPRGCLPTGDRGTGLEVALVRSDTIANIQDNLSNMAATAPFNDAALL
 NMYKSNKPEAGDLARALEEPLFCSCAGYAVLVIGIDPHSDNIMERSGOLFIDFEGH
 FLNFKTKFGINERVDFILTYDFVHHYIQGKKNSKFKFRFGYCGRAATYIIRHDL
 LFLHRLMRAAGLPELSCSKDIQYLKDSLALCKTEDEALAKHFRVKNENALRSWKKK
 VNMIAHVSNDNQ"

```

Query Match      45.9%  Score 26.6;  DB 10;  Length 3668;
Best Local Similarity 74.3%  Pred. No. 22;
Matches 26;  Conservative 5;  Mismatches 4;  Indels 0;  Gaps 0;

OY      24  GATGATGGCCACATGCTGTTTCATATMGATTTTGGCCA  58
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
db      2903  GAGAGTGGGCGAGCTGTTCCACATTTGATTTTGGCCA  2937

```

```

RESULT 12
HSU57843
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS

```

HSU57843 3525 bp mRNA PRI 10-MAY-1997
 Human phosphatidylinositol 3-kinase delta catalytic subunit mRNA,
 complete cds.
 U57843
 92076750
 U57843.1 GI:2076750

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 3525)
 Mahlum,C.E., Becker,K.P. and Morris,A.J.
 H.sapiens mRNA for phosphoinositide 3-kinase delta catalytic
 subunit
 unpublised
 2 (bases 1 to 3525)
 Morris,A.J.
 Direct Submission
 Submitted (09-MAY-1996) Andrew J. Morris, Pharmacology, SUNY-Stony
 Brook, BST, Level 7 Room 168, SUNY-HSC, Stony Brook, NY 11794-8651,
 USA

location/Qualifiers
 1. 3525
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 116. 3322
 /codon_start=1
 /product="phosphatidylinositol 3-kinase delta catalytic

supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites. CDS are numbered using the following system eg SPBC25H2.01c.SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid c1672 is overlapped by cosmid c1183 at its 3' and by cosmid c962 at its 5'.

FEATURES

```

source
    location,Qualifiers
    1..36000
    /organism="Schizosaccharomyces pombe"
    /strain="972h"
    /db_xref="taxon:4896"
    /chromosome="III"
    /clone="cosmid C1672"
    /map="III"
    /complement(1..1848)
    /note="nominal overlap with c962 S. pombe chromosome 3"
    1579..2789
    /gene="SPCC1672.01"
    join(1579..1610,1727..1853,2028..2789)
    /gene="SPCC1672.01"
    /note="SPCC1672.01, len:306aa, similarity: yeast, YFR025C, HIS9_YEAST, histidinol-phosphatase, (335aa), fasta scores, opt:306, E():0, (33.3% identity in 330 aa overlap)"
    /codon_start=1
    /product="histidinol-phosphatase"
    /protein_id="CA20439.1"
    /db_xref="PID:e1316727"
    /db_xref="GI:3451302"
    /db_xref="GI:3451302"
    /translation="MPISSESHSGORCLNAQGLKEDVIOGATQCGRSFPEHPRD
RVEDLPRELLHOPEDLRTFPEYVNEARLKNQNGDQINILGATETIRPESVLL
KSLNTRYNDIVGVSHVNSIPIDSPELWQKALQHVNNBQGLDIFDEIQYDLMQ
RHAPLVIGFEDLICLEFADPAKVEFNKSNVWELLQRNLIKAVSYGCIETISAFRK
GKATVAPQORLELMEVQGAQLSDSPSHGPGVQVGLNYHLAKSYDLKCGTSLCMTEK
GPDNGTVKNTVDMVMSKFGTNGITNT"
    1611..1616
    /gene="SPCC1672.01"
    /note="gtagtg, splice donor sequence"
    1711..1726
    /gene="SPCC1672.01"
    /note="ctaaactttatggaag, splice branch and acceptor"
    1854..1859
    /gene="SPCC1672.01"
    /note="gtaagt, splice donor sequence"
    2009..2027
    /gene="SPCC1672.01"
    /note="ctaaactatgattcattag, splice branch and acceptor"
    /complement(3865..4629)
    /gene="sapi"
    complement(3865..4629)
    /gene="sapi"
    /note="SPCC1672.02c, len:254aa"
    /codon_start=1
    /product="switch-activating protein 1"
    /protein_id="CA20440.1"
    /db_xref="PID:e1316728"
    /db_xref="PID:g3451303"
    /db_xref="GI:3451303"
    /translation="MEAPKMLKSYRKKNASLSPSSSPAKAQRTHLSLEKIKRLMY
VHKHLEINDRKTSFPAKATRGVDEGALITBESRCNOISIMRYAVQALAHQPS
MKTTPEDDLDDCEMKARLSELQOYRKRFVGRKKCCNDSEINERKLLLEDOQNV
DKLVAKVFLSLHLDNEERKLMQVNAKMEVLAENKRIQOOLLHDNDLLSKLTPPSA
PGVNGMTNMGANMGANNAIRGGLHSSISIPNLGDH"
    complement(7403..8986)
    /gene="SPCC1672.03c"

```

```

CDS
    complement(7403..8986)
    /gene="SPCC1672.03c"
    /note="SPCC1672.03c, len:527aa, similarity: YDL238C, YD3C_YEAST, hypothetical 55.2 kd protein, (489aa), fasta scores, opt:1131, E():0, (43.6% identity in 473 aa overlap)"
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="CA20441.1"
    /db_xref="PID:e1316729"
    /db_xref="PID:g3451304"
    /db_xref="GI:3451304"
    /translation="WDGRCDFVGVCKLIHPISGIEETTDPAVGVYKNKIVLEKMT
PKTLEAKSHLLKATITHKLPIDQFNEPGLDITHAPQYPNSIGIDVPILOMLEK
ITFPLESSLADLEEAROVYKRVERTLSNGTFASVSTLHPITSLALAEICYSYGR
AIVGCKNNMNSLPHYCCERKSAESSLEAROLIYSMSIDPKREMYTPITRPAFCF
EDLISGCELEAKENMLPIOTHESENTSEIELEKLPKRSYADVDYKLLTPOTIL
MAHILDEBEITELTKRSSGISHCPTSILASGLANRKILDSGINVGLDVGSGY
APSLIILARHAMTSRSLSYVIGDPKVLDSLILXYLAQGAPEVSRGDVGSRAG
KYWDLIYDLSAEHRSYDIFERDTPVPLSKWFTSDRNLAQVAVNGRLVSGFEMK
ANLKSTPLTNGVTSISGHQVPELTOAHLPRTOCDVDPSCCGGCKEESCRTENC
KGAYPANATVVEEDSGMS"
    complement(10212..10584)
    /gene="SPCC1672.04c"
    complement(join(10212..10465,10533..10584))
    /gene="SPCC1672.04c"
    /note="SPCC1672.04c, len:101aa"
    /codon_start=1
    /product="very hypothetical protein"
    /protein_id="CA20442.1"
    /db_xref="PID:e1316730"
    /db_xref="PID:g3451305"
    /db_xref="GI:3451305"
    /translation="MTTRPEPERSFPDLHGEECTYVKNQYKLECIKRENOEBCRL
AKKYLCRMDTGLFGKDMKNIGFGDENASTLSSSNDGNMNSSSSDNKTKGE"
    complement(10466..10483)
    /gene="SPCC1672.04c"
    /note="ctaaactatgctgcgag, splice branch and acceptor"
    complement(10527..10532)
    /gene="SPCC1672.04c"
    /note="gtagtg, splice donor sequence"
    complement(11755..12960)
    /gene="SPCC1672.05c"
    complement(11755..12960)
    /gene="SPCC1672.05c"
    /note="SPCC1672.05c, len:401aa"
    /codon_start=1
    /product="tyrosyl-tRNA synthetase cytoplasmic EC 6.1.1.1"
    /protein_id="CA20443.1"
    /db_xref="PID:e1316731"
    /db_xref="PID:g3451306"
    /db_xref="GI:3451306"
    /translation="MSLTPDERKYLITRNLEQVLAGKMKIREILNERDISLYWSAPTG
RPHCGYFVPMKMLADFLKAGVHYVLADVAFLDNKMAPHELVQHVRYEIIYKSI
LKSINPIEKLREVISSGYOLKEDYTLNPLFASITRNPRBAGAEVQVQVAPSLAS
SLIYRGMALDEQVLYGVYQCGVDQKRIIPNAEVLVLYCYKKRGHLMNVPYELAG
GMSSSAANAKRIDLDDPKTVNKRIRAFCEPGEVNEGCIAFKIVSFPAEMLRGE
TEFVLDREDFQEGRAVFETVEVERKAYDLSLQDLGLGESSVNTLLAAYVQDLKO
YPDIDSLIKAAVPPDKDLKRAMKOKQDKRKNKAGTTNASTIAESGAPESTQTA
SEFKRLND"
    complement(11875..12930)
    /gene="SPCC1672.05c"
    /note="Pfam match to entry PF00579 tRNA-synt_1b, tRNA synthetases class I (Tyr and Tyr)"
    complement(13222..15984)
    /gene="SPCC1672.06c"
    complement(13222..15984)
    /gene="SPCC1672.06c"
    /note="SPCC1672.06c, unknown, len:920aa, similarity: S. cerevisiae, YLR410w, Q06885, chromosome xii cosmid 9931.7, (1146aa), fasta scores, opt:2648, E():0, (52.5% identity in 962 aa overlap)"
    /codon_start=1

```

/product="conserved hypothetical protein"
 /protein.id="CAA02444.1"
 /db.xref="PID:ej316732"
 /db.xref="PID:g3451307"
 /db.xref="GI:3451307"
 /db.xref="SPRFBHL:074429"
 /translation="MIONASHLSTIDTSESTRASFVSTYPTKRVNYGICAMDAK
 VSKRKNILIRLITAECEFEVLEVDGNMLIDLEVENNPAADYLCISGSGFPLKKEK
 LERKPCPVNDVAFQELMDRRLVYLIDAIRSPSTLIRCDGGPIMNVLEKRA
 RKPELEITETVPEYMLDEDTLSVSGKIIEKRVYKPGGDHNYITFPRVGGG
 RKTLEKTAANSSTYDPLCAPRGEGSIKEKRVNMDADVYVGVGHYSHAETK
 PVVDGIVRNPDHGELEITLISEEKNAKSLIAFEGVCPFLPILVSCOSYIDV
 NGSEVLENDNDYDNNARLIKQFHAERHRRVRSVQVNNPPRSEMAKSLSL
 IYLSHARLEPRKGFSEFSDPVKTLQSHTEVILRNQLNLSGAALNPLKCE
 INKTLQRLALEKFDLPKRVKLVKQVPEKTLIKLLIIMWGEFPHSNARSKD
 GDSHRAQDLYMNDKTLKNDYVYSRRRYSNEMAFLEQETIPSDLKVRKDL
 GEENRAQDLYMDKTKHKLSLRIVGDTARKEDTIPENMRKCPQGVQLKILHAY
 ENNTILGPEVQVQSWCCNENPALFRWMEVTSFEDSCAPKSVSEVLYDLTKL

Query Match	45.5%	Score 26.47	DB 7	Length 36000
Best Local Similarity	58.6%	Pred. No. 29		
Matches 34, Conservative	6	Mismatches 18	Indels 0	Gaps 0

Q7 1 AARATGAGATGGATHTTTCACACGATATGGCCARCTGTTTTCATATWGAATTTGGCCA 58
|| || | : : || : | | | | | : || : | |||
Db 19047 AACATTGAAGGTTGCTCCATAAGCTAATTTCCACCTGGTCACACAGCGTTGTGGCA 18990

Db 19047 AACATTGAAGGTTCTCCATAAGCTAATTCACCACCTGGTCCACATGCGTTGTGCCA 18990

RESULT 15

DEFINITION *Caenorhabditis elegans* cosmid T21E12.
ACCESSION U00440

VERSION T80440.1 GI:1703596
KEYWORDS

ORGANISM	Caenorhabditis elegans
Eukaryota: Metazoa: Nematoda: Secernentea	

REFERENCE	1 (bases 1 to 35820)
WILSON R	Ainscough B
AUTHORS	Andersen K
	Baynes C
	Perle M

Coulson, A., Craxton, M., Bear, S., Du, Z., Durbin, R., Favellio, A., Fyfe, J., Gardner, A., Green, P., Hawkins, T., Hillier, L., Tier, M.

O'Callaghan, M., Parsons, J., Percy, C., Rifkin, J., Roopra, A.,
Latreille, P., Lightning, J., Lloyd, C., McMurphy, A., Mortimore, B.,

Scadden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, B., Watson, A., Weinstein, I.,

LINE 22 MD OF CONGRUOUS NUCLEOLAR SEQUENCE FROM CHROMOSOME III OF C. elegans

REFERENCE	2 (bases 1 to 35820)
MEDLINE	34130/18

THE
JOURNAL
Unpublished (1996)

TITLE	Direct Submission
-------	-------------------

Genome Sequencing Center

Sanger Centre, Hinxton Hall

NOTICE: This sequence may not be the entire insert of this clone.

COMMENT Submitted by:

St. Louis, MO 63110, USA, and

e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

CC These sequences were transformed into Schizosaccharomyces pombe cells
 CC in an embodiment of the invention. In the presence of thiamine the
 CC promoter is inactive and the cells carrying the PKC plasmids grow as
 CC the parental strain. In the absence of thiamine the nmt promoter
 CC functions and the PKC is induced. PKC activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC ptdins, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 CC Sequence 38 BP; 7 A; 6 C; 8 G; 11 T;

Query Match 56.9%; Score 33; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 0.00091;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GATGATGGCCARCTGTTTCATATGATGATTTGGCCA 58
 DB 1 GATGATGGCCARCTGTTTCATATGATGATTTGGCCA 35

RESULT 3
 ID 051155 standard; CDNA; 3207 BP.
 AC 051155;
 DT 12-APR-1994 (first entry)
 DE p110 CDNA.
 KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; propylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ss.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /*tag= a
 FT /note= "PI3- kinase p110"
 PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Vollinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 DR P-PSDB; R43341.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Claim 7; Fig 9; 146pp; English.
 CC Southern blot analysis was performed using a bovine CDNA probe contg.
 CC a fragment of a PI3-kinase-encoding sequence and human CDNA isolated
 CC from a CDNA library constructed from mRNA isolated from the human
 CC cell line K562. Positive clones were sequenced to give the human
 CC PI3 kinase p110 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p110 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to
 CC provide a protein with PI3 kinase activity, and is useful for
 CC screening for (ant)agonists of PI3 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC propylaxis or therapy. Platelet or neutrophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051156, 059012-23 and 057522-3.
 CC Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match 52.1%; Score 30.2; DB 1; Length 3207;
 Best Local Similarity 78.4%; Pred. No. 0.043;
 Matches 29; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 22 AGCATGATGGCCARCTGTTTCATATGATGATTTGGCCA 58
 DB 2771 AGCATGATGGCCARCTGTTTCATATGATGATTTGGCCA 2807

RESULT 4
 ID 057012 standard; CDNA to mRNA; 3498 BP.
 AC 057012;
 DT 31-AUG-1994 (first entry)
 DE ptdins 3-kinase 110 kD catalytic subunit CDNA.
 KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KW ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Bos taurus.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /*tag= a
 FT /product= p110

PN WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-01654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 DR P-PSDB; R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure; Fig 1; 71pp; English.
 CC This sequence encodes the 110 kD catalytic subunit of the phosphatidylinositol (ptdins) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nmt promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the ptdins
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nmt promoter functions and the ptdins 3-kinase
 CC catalytic subunit is induced. Ptdins activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 CC Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 52.1%; Score 30.2; DB 1; Length 3498;
 Best Local Similarity 78.4%; Pred. No. 0.045;
 Matches 29; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 22 AGCATGATGGCCARCTGTTTCATATGATGATTTGGCCA 58
 DB 2771 AGCATGATGGCCARCTGTTTCATATGATGATTTGGCCA 2807

RESULT 5
 ID 020959 standard; DNA; 1696 BP.
 AC 020959;
 DT 12-MAY-1992 (first entry)
 DE L-dextranase glucose-6-phosphate dehydrogenase coding sequence.
 KW G6P-DH; glycolysis; enzyme; DSM 20187; thermal stability; ss.
 OS Leuconostoc dextranans;
 FH Key Location/Qualifiers
 FT cds 1..122
 FT promoter
 FT /*tag= a
 FT /*tag= b
 FT /product= G6P-DH
 PN EP-469523-A.
 PD 05-FEB-1992.
 PF 29-JUL-1991; 112739.
 PR 30-JUL-1990; DE-024158.
 PA (BOE) BOEHRINGER MANNHEIM GMBH.
 PI Jarsch M, Lang G;

```

DR WPI; 92-042914/06.
DR P-FSD8; R20785.
PT Recombinant glucose-6-phosphate dehydrogenase - derived from
PT Leuconostoc dextranucis, for assaying glucose-6-phosphate, with
P5 Improved temp. stability
P5 Claim 2; Page 9; 13pp; German.
CC The G6P-DH gene was isolated from L.dextranucis genomic DNA
CC fragments using a probe corresponding to part of the enzyme from
CC transformed with plasmid pUC-G6P-DH.1.8 (containing the G6P-DH gene
CC under control of the native Leuconostoc protein) was recovered as
CC crystals. The enzyme was found to have lower Km, lower inhibition by
CC glycerol and Mg ions and better temp. stability than the natural
CC enzyme. The enzyme also shows no reaction with glucose. See also
CC Q20360.
SQ Sequence 1696 BP; 523 A; 310 C; 362 G; 501 T;

Query Match 51.7%; Score 30; DB 1; Length 1696;
Best Local Similarity 62.0%; Pred. No. 0.04;
Matches 31; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 1 AARATGATGATGATHTTCAACAGATGATGCCACACTGTTCAATACATGAY 50
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 607 AAGAGCATTTGGAAATGCAATGCAATTTGATGATGACCACTGTCGTATGAC 656

RESULT 6
T30577
ID T30577 standard; DNA; 1455 BP.
AC T30577:
DT 14-SEP-1996 (first entry)
DE Leuconostoc dextranucis glucose-6-phosphate dehydrogenase gene.
KW Glucose-6-phosphate dehydrogenase; G6PDH; Immunooassay; ss.
OS Leuconostoc dextranucis strain ATCC 19255.
PN M09424559-A2.
PD 27-OCT-1994.
PE 07-APR-1994; U03437.
PR 08-APR-1993; U5-044857.
PA (BEHM ) BEHRINGERWERKE AG.
PA (GEWV ) GENENCOR INT INC.
PA (SYNT ) SYNTex USA INC.
SI Barnett CC, Becker MJ, Bott RR, Caldwell RM, Goodman TC;
SI Jakobovits EB, Levy MJ, Sillen JL, Ullman EF;
DR WPI; 95-006211/01.
P-P5DB; R95964.
PT New mutant glucose-6-phosphate dehydrogenase enzymes - used partic.
PT to form conjugates for use in homogeneous immunoassays for analytes
PS Disclosure; Page 62-63; 122pp; English.
CC G6PDH genes (T30574-77) respectively code for the glucose-6-phosphate
CC dehydrogenases (R95961-64) of Leuconostoc mesenteroides ATCC 12291,
CC Leuconostoc citrum NCIMB 3351, Leuconostoc lactis NCMO 546 and
CC Leuconostoc dextranucis ATCC 19255. The enzymes may be mutated to
CC increase stability, or to improve or modulate activity, pref. by
CC deletion or substn. of 1 or more lysine residues, or introduction of
CC 1 or more cysteine residues. The mutant enzymes are prepd. by mutating
CC the G6PDH gene and expression in host cells. They are used partic. to
CC form conjugates for use in homogeneous immunoassays for analytes.
SQ Sequence 1455 BP; 452 A; 275 C; 318 G; 410 T;

Query Match 51.7%; Score 30; DB 1; Length 1455;
Best Local Similarity 62.0%; Pred. No. 0.04;
Matches 31; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 1 AARATGATGATGATHTTCAACAGATGATGCCACACTGTTCAATACATGAY 50
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 482 AAGAGCATTTGGAAATGCAATGCAATTTGATGATGACCACTGTCGTATGAC 531

RESULT 7
ID T305156 standard; cDNA; 3412 BP.

```

```

AC Q5156-1994 (first entry)
DE Human p110 cDNA.
KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ds.
OS Human.
FH Key Location/Qualifiers
FT cds 1..3207 /tag= a
FT /note= "PI3-kinase p110"
PD WO9321328-A.
PF 28-OCT-1993.
PF 13-APR-1993; G00761.
PA (UDW-) LODWIC INST CANCER RES.
PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI Parker J, Volinia S, Waterfield MD;
DR WP1_9335178/44.
P-PEDB: R43342.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Claim 7; Fig 16; 146bp: English.
CC Southern blot analysis was performed using a bovine cDNA probe contg.
CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
CC from a cDNA library constructed from mRNA isolated from the human
CC cell line K51A. Positive clones were sequenced to give the human
CC PI3 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with PI3 kinase activity, and is useful for
CC screening for (ant)agonists of PI3 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also Q5155 and Q57522-3.
SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match          49.3%; Score 28.6; DB 1; Length 3412;
Best Local Similarity 75.7%; Prcd No. 0.17;
Matches 28; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY      22 ACAGATGATGGCCARCTGTTCAYATWAGATTGGCCA 58
        | | ||||| ||:|||||:::||:||||||| ||
Db       2771 AGACGATGCACAACACTGTTTCAATAGATTGGACA 2807

RESULT      8
X13228
AC X13228 standard; DNA: 4956 BP.
DE 19-MAR-1999 (first entry)
KW Enterococcus faecalis genome config SEQ ID NO:291.
KW Enterococcus faecalis; config; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN WO950555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PU, Kunsch CA;
PI WPI: 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS - Claim 1; Page 1327-1330; 2084pp; English.
```


RESULT 10
X20017/c

RESULT	11
ID	024526
AC	Q24526 standard; DNA; 1963 BP.
DT	06-NOV-1992 (first entry)
DE	Glc6pP genome.
KW	Glucose-6-phosphate dehydrogenase; NAD+; NADP+; enzyme mechanism; site directed mutagenesis; ss.
OS	Leuconostoc mesenteroides.
EH	Key
FT	Location/Qualifiers
FT	cds
FT	215..1669
FT	/*tag= a
PN	M09207078-A.
PD	30-APR-1992.
PF	15-OCT-1991; 007715.
PR	12-OCT-1990; US-596867.
PA	(UYSY-) UNIV SYRACUSE.
PI	Lee WT, Levy HR;
DR	WPI, 92-167158/20.
DR	P-PSDB; R23853.
PT	Isolation of Glc6pP genome of <i>L. mesenteroides</i> - prepd. by PCR
PT	using appropriate 3' and 5' primers and cloned into pUC19 for
PS	expression in <i>E. coli</i>
PS	Disclosure; Page 10; 22pp; English.
CC	The sequence given encodes the glucose-6-phosphate dehydrogenase
CC	gene (Glc6pP) from <i>L. mesenteroides</i> . The gene product of this
CC	sequence is a dimer with identical subunits of mol. wt. 54,800. It
CC	can utilize either NAD+ or NADP+ as coenzyme in the oxidation of
CC	glucose-6-phosphate. The kinetic mechanism differs depending on
CC	whether NAD+ or NADP+ is used in the reaction. Binding of NAD+
CC	


```
CC produces a large conformational change in the enzyme.
CC This sequence can be cloned into E. coli in a suitable expression
CC vector such that large amounts of Glc6PD can be produced economically.
CC The elucidation of the Glc6PD sequence allows sequence comparison
CC with the equivalent gene in other species eg. Drosophila melanogaster
CC and rat. Such information can be used to design site directed
CC mutagenesis experiments to help determine the enzyme mechanism.
SO Sequence 1963 BP; 599 A; 403 C; 394 G; 567 T;

Query Match      46.2%; Score 26.8; DB 1; Length 1963;
Best Local Similarity 58.0%; Pred. No. 0.67;
Matches 29; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

OY 1 AARATGAYTGGATTTCATACAGATGATGCCACTGTTCATCATWAV 50
   ||||| :|||| :|| :||||| ||||| ||:|| ||:|||
Nh 496 AAAATGACTTGGAACACGATTGTGATTAACCACTATTCGTAFTGAC 745

RESULT 12
T30574 T30574 standard; DNA; 1458 BP.
ID T30574
AC T30574;
DE 14-SEP-1996 (first entry)
Leucostoc mesenteroides glucose-6-phosphate dehydrogenase gene.
KW Glucose-6-phosphate dehydrogenase; G6PDH; Immunoassay; ss.
OS Leucostoc mesenteroides strain ATCC 12291.
PN W09424559-A2.
PD 27-OCT-1994.
PF 07-APR-1994; U03437
PR 08-APR-1993; US-044857.
PA (BEHM ) BEHRINGERKE AG.
PA (GENV ) GENENCOR INT INC.
PI (SYNT ) SYNTEX USA INC.
PI Barnett CC, Becker MJ, Bott RR, Caldwell RM, Goodman TC;
PI Jacobowitz EB, Levy MJ, Sillen JL, Ullman EF;
DR WPt: 95-006211/01.
P-PSTD: R95961.
PT New mutant glucose-6-phosphate dehydrogenase enzymes - used partic.
PT to form conjugates for use in homogeneous immunoassays for analytes
PS Disclosure: Page 54-55; 12pp; English.
CC G6PDH genes (T30574-77) respectively code for the glucose-6-phosphate
CC dehydrogenases (R95961-64) of Leucostoc mesenteroides ATCC 12291,
CC Leucostoc citreum NCIMB 3351, Leucostoc lactis NCD0 546 and
CC Leucostoc dextranlicum ATCC 19259. The enzymes may be mutated to
CC increase stability, or to improve or modulate activity, pref. by
CC deletion or substn. of 1 or more lysine residues, or introduction of
CC 1 or more cysteine residues. The mutant enzymes are prepd. by mutating
CC the G6PDH gene and expression in host cells. They are used partic. to
CC form conjugates for use in homogeneous immunoassays for analytes.
SO Sequence 1458 BP; 430 A; 310 C; 313 G; 405 T;

Query Match      46.2%; Score 26.8; DB 1; Length 1458;
Best Local Similarity 58.0%; Pred. No. 0.61;
Matches 29; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

OY 1 AARATGAYTGGATTTCATACAGATGATGCCACTGTTCATCATWAV 50
   ||||| :|||| :|| :||||| ||||| ||:|| ||:|||
Db 482 AAAATGACTTGGAACACGATTGTGATTAACCACTATTCGTAFTGAC 531

RESULT 13
T30576 T30576 standard; DNA; 1461 BP.
ID T30576
AC T30576;
DE 14-SEP-1996 (first entry)
Leucostoc lactis glucose-6-phosphate dehydrogenase gene.
KW Glucose-6-phosphate dehydrogenase; G6PDH; Immunoassay; ss.
OS Leucostoc lactis strain NCD0 546.
PI Key Location/Qualifiers
FH misc_difference 187..213
FT
```

```

FT      /note- "bases n at positions 187-213 are not
FT      Identified in the specification"
FT      misc_difference 1051..1089
FT      /tag= b
FT      /note- "bases n at positions 1051-1089 are not
FT      identified in the specification"
FT      misc_difference 1171..1185
FT      /tag= c
FT      /note- "bases n at positions 1171-1185 are not
FT      identified in the specification"

PN      WO924559-A2.
PD      27-OCT-1994.
PP      07-APR-1994; U03437.
PR      08-APR-1993; US-044857.
PA      (BEHM ) BEHRINGERMERE AG.
PA      (GENV ) GENENCOR INT INC.
PI      Barnett CC, Becker MJ, Bott RR, Caldwell RM, Goodman TC,
PI      Jakobovits EB, Levy MJ, Sillen JL, Ullman EF;
DR      P-PDSB; R95963.
PT      New mutant glucose-6-phosphate dehydrogenase enzymes - used partic.
PT      to form conjugates for use in homogeneous immunoassays for analytes
PS      Disclosure; Page 60; 122pp; English.
CC      G6PDH genes (T30574-77) respectively code for the glucose-6-phosphate
CC      dehydrogenases (R95961-64) of Leucosotoc mesenteroides ATCC 12291,
CC      Leucosotoc citreum NCIMB 3351, Leucosotoc lactis NCO 546 and
CC      Leucosotoc dextranum ATCC 19255. The enzymes may be mutated to
CC      increase stability, or to improve or modulate activity, pref. by
CC      deletion or substn. of 1 or more lysine residues, or introduction of
CC      1 or more cysteine residues. The mutant enzymes are prepd. by mutating
CC      the G6PDH gene and expression in host cells. They are used partic. to
CC      form conjugates for use in homogeneous immunoassays for analytes.
SQ      Sequence 1461 BP; 395 A; 284 C; 324 G; 377 T;

Query Match          46.2%; Score 26.8; DB 1; Length 1461;
Best Local Similarity 58.0%; Pred. No. 0.61;
Matches 29; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

OY      1 AARAGCATGTGATTTTCATACGCATGTGGCCACTCTTTCATATTNAY 50
DB      485 AAAGAATTGGMAAACCGCTTGATGACCAATTGTCGATTGAC 534

RESULT 14
VI6533 standard; cDNA; 3387 BP.
VI6533 AC VI6533;
DT      22-JUN-1998 (first entry)
DE      cDNA encoding mammalian novel class I p13 kinase designated p110-delta.
KW      Autophosphorylate; PI3; phosphatidylinositol 3-hydroxy; kinase;
KW      activity; p110-delta; melanoma; lipid kinase; metastase;
KW      phosphoinositide specificity; regulation; diagnosis; motility;
KW      cancer cell; control; ss.
OS      Homo sapiens.
FH      Key Location/Qualifiers
ET      CDS 1..3387
ET      CD5 /tag= a
PN      WO9746688-A1.
PD      11-DEC-1997.
PF      30-MAY-1997; G01471.
PR      01-JUN-1996; GB-011460.
PA      (LUDW-) LUDWIG INST CANCER RES.
PI      Vanhasbroeck B, Waterfield MD;
DR      WPI; 98-042196/04.
DR      P-PDSB; WA6625.
PT      Auto:phosphorylating peptide with phosphatidyl.inositol
PT      3-kinase-like activity designated pl10delta - useful for
PT      controlling cell motility, particularly of metastatic cancer cells
PS      Disclosure; Fig 9; 72pp; English.
CC      The present sequence encodes a novel autophosphorylating protein
CC      that has PI3 (phosphatidylinositol 3-hydroxy) kinase activity, and

```

CC is designated p110-delta. The protein is expressed selectively in
 CC white blood cells and melanomas. The protein is a lipid kinase with
 CC broad phosphoinositide specificity and specific tissue localization,
 CC possibly involved in regulation of melanoma metastases (it is not
 CC expressed in normal melanocytes). Fragments of the present
 CC sequence are used to detect tissue-specific expression, and a
 CC similar analysis can be done at the protein level using antibodies in
 CC standard immunoassays. These assays are particularly used for diagnosing
 CC and predicting motility/invasiveness of metastatic cancer cells. The
 CC protein can be used in human or veterinary medicine for controlling
 CC motility of cells, where the protein increases motility while antisense
 CC sequences are used to reduce it.
 CC Sequence 3387 BP: 711 A; 1044 C; 1001 G; 631 T;

Query Match 45.9%; Score 26.6; DB 1; Length 3387;
 Best Local Similarity 74.3%; Pred. No. 0.95;
 Matches 26; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 24 GATGATGGCCARCTGTTTCATATGATTTGGCCA 58
 DB 2707 GAGAGTGGCAGCTGTTCCACATTTGTTGGCCA 2741

RESULT 15
 ID V31340 standard: cDNA: 5220 BP.
 AC V31340;
 DT 12-OCT-1998 (first entry)
 DE Human phosphatidylinositol 3-kinase p110 delta subunit cDNA.
 KW phosphatidylinositol 3-kinase; p110 delta; human; immune system;
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 196..3350
 FT /tag= a
 PN WO9823760-A1.
 PD 04-JUN-1998.
 PF 25-NOV-1997: U21655.
 PR 25-NOV-1995: US-777405.
 PA (ICOS-) ICOS CORP.
 PI Chantry DH, Hoeckstra MF, Holtzman DA;
 DR MPI: 98-322736/28.
 DR P-PSDB: W58570.
 PT New phosphatidylinositol 3-kinase catalytic subunit - used to
 PT develop products for modulating kinase activity in immune system
 PT signalling and in carcinogenesis
 PS Claim 4: Page 27-33; 53pp: English.
 CC This full-length composite cDNA encodes the p110 delta catalytic
 CC subunit (see W58570) of human phosphatidylinositol 3-kinase. It
 CC was assembled from clone #249 obtained by PCR amplification (see
 CC W31341-44) of human peripheral blood mononuclear cell cDNA, clone
 CC #0928 obtained by screening a human macrophage cDNA library, and
 CC further clones obtained by RACE and PCR (see V31345-50). The
 CC following are claimed: (1) a purified and isolated polynucleotide
 CC (PN) encoding p110 delta; (2) a vector comprising a DNA as in (1);
 CC (3) a host cell stably transformed or transfected with a DNA as in
 CC (1); (4) PN encoding a lipid kinase, and hybridizing to PN having
 CC the 5220 bp sequence; (5) a purified and isolated p110 delta
 CC polypeptide as in (4); (6) an antibody specifically immunoreactive
 CC with p110 delta; (7) a hybridoma cell line (especially 208F
 CC (HB 12200) producing a monoclonal antibody as in (6); and (8) a
 CC humanized antibody as in (6). p110 delta has kinase activity and
 CC may play a role in PI 3-kinase mediated signalling in the immune
 CC system and in carcinogenesis. The products can be used to develop
 CC agents that modulating p110 delta kinase activity and to develop
 CC diagnostic reagents (claimed). They may also be used for detection
 CC and diagnosis of p110 delta in a biological sample.
 SQ Sequence 5220 BP: 1120 A; 1525 C; 1483 G; 1092 T;

Query Match 45.9%; Score 26.6; DB 1; Length 5220;
 Best Local Similarity 74.3%; Pred. No. 1.1;

Matches 26; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 24 GATGATGGCCARCTGTTTCATATGATTTGGCCA 58
 DB 2902 GAGAGTGGCAGCTGTTCCACATTTGTTGGCCA 2936

Search completed: September 13, 1999, 16:21:48
 Job time: 4008 sec

RESULT	1
AI324274	
LOCUS	
DEFINITION	453 bp mRNA EST 23-DEC-1998
	mq2cc02.y1 Stragene mouse melanoma (#937312) Mms musculus cDNA
	clone IMAGE:584758 5' similar to SW:PIA.MOUSE P42337
	PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
	mRNA sequence.
ACCESSION	AI324274

AUTHORS Stewart,G.J., Savioz,A. and Davies,R.W.
TITLE Sequence analysis of 497 mouse brain ESTs expressed in the
substantia nigra
JOURNAL Genomics 39, 147-153 (1997)
COMMENT On Nov 22, 1995 this sequence version replaced gi:1070910.

Contact: Davies, R.W.
Robertson Laboratory of Biotechnology
Institute of Biomedical and Life Sciences, Division of Molecular
Genetics
Robertson Building, 54 Dumbarton Road, Glasgow G11 6NU
Tel: 44 141 330 5102
Fax: 44 141 330 5102/4878
Email: gbgar1eudcf.gla.ac.uk
Seq primer: T7/T3alpha
High quality sequence, stop: 528.

names	Location/Qualifiers
source	1. .528

BASE COUNT	158 a	85 c	121 g	164 t
ORIGIN				

Query Match	43.1%	Score 25;	DB 26;	Length 528;
Best Local Similarity	56.9%	Pred. NO. 32;		
Matches 29; Conservative	8;	Mismatches 14;	Indels 0;	Gaps 0

QY 4 ATGAYTGGATHTTTCAYACGATGATGGCCARCTGTTTCATWGYTTTG 54
 ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
 Db 147 ATGCATTCATCTTCTCACCCAGATGGACATCTGTGCCACACTGATTTTG 97

RESULT	4
LOCUS	AA174738
DEFINITION	AA174738 476 bp mRNA EST 16-FEB-1997
	mt12L1.r1 Soares mouse 3DNBS P4338 musculus cDNA clone IMAGE:620876
	5' similar to SW:PI1B.HUMAN P4338 PHOSPHATIDYLINOSITOL 3-KINASE
	CATALYTIC SUBUNIT, BETA ISOFORM ;, mRNA sequence.

ACCESSION	AA174738
NID	91755862
VERSION	AA174738.1
KEYWORDS	GI:1755862
SOURCE	EST.
ORGANISM	house mouse. Mus musculus

REFERENCE
1 (bases 1 to 476)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.

TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Apr 14, 1993 this sequence version replaced g1:693229

Contact: Maura M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63106
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:381700
 Seq primer: -28M13 rev2 from Amerham
 High quality sequence stop: 475.
 Location/Qualifiers
 1..476

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="5 p15.2-p15.1"
/clone="IMAGE:620876"
/clone_lib="Soares mouse 3MBMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTATCCATCTGAAAGTGGGACGCCGCGCTGTATTTTATTTTATTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M. Fatima Bonaldo."

```

Query Match	43.1%	Score 25	DB 29	Length 476
Best Local Similarity	71.4%	Pred. NO. 32		
Matches 25	Conservative 5	Mismatches 0	Indels 0	Gaps 0

```
QY      24 GATGATGGCCACCTGTTCAYATWAGYTTTGCCA 58  
        || || |::|::|::|::|::|::|  
Db     37 GAGAGTGGCAGACTTCACATTGATTTTGSCCA 71
```

RESULT	5
R12466	
LOCUS	
DEFINITION	
	R12466 328 bp mRNA EST 12-APR-1995
	yf6bh7.r1 Soares infant brain IN1B Homo sapiens cDNA clone
	IMAGE:26444 5' similar to SP:P100 BOVIN P32871 PHOSPHATIDYLINOSITOL
	3-KINASE CATALYTIC SUBUNIT ;, mRNA sequence.

ACCESSION	R12466
NID	9765542
VERSION	R12466.1
KEYWORDS	GI:765542
SOURCE	EST.
ORGANISM	human.
	Homo sapiens

REFERENCE 1 (bases 1 to 328)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

TITLE	COMMENT
The WashU-Merck EST Project Unpublished (1995)	

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 2060

High quality sequence stops: 256 Source: IMAGE Consortium, L1NN1
This clone is available royalty-free through L1NN1; contact the
IMAGE Consortium (Info@image.linn.gov) for further information.

CDNA Library Preparation: Matthew Clark, cDNA Library Arrayed by: Matthew Clark, DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: Genome Systems, St. Louis Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com)
Resource center: umir.mdb.tu-berlin.de Berlin, Germany (web address: www.rnd.de)

TITLE
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dime, D., Feng, D.-F., Ferrite, A., Fischer, C., Hastings, G.A., He, M.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hung, J., Li, H., Melsner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C.
JOURNAL
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
MEDLINE
Nature 377 (6547 Suppl), 3-174 (1995)
COMMENT
On May 5, 1995 this sequence version replaced g1:798292.

FEATURES
source
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
1..243
/organism="Homo sapiens"
/db_xref="ATCC (inhost):188107"
/db_xref="taxon:9606"
/map="19"
/clone_1lb="Thymus II"
/sex="male"
/dev_stage="adult, 19 yrs"
/note="Organ: thymus; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
/score: Site_2: XhoI"

BASE COUNT
74 a 63 c 37 g 69 t
ORIGIN

Query Match
Best Local Similarity 52.8%; Score 23.4; DB 33; Length 243;
Matches 28; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY
3 RATGAYTGATHTTTCATACATGATGCGCATCTTTCATATGATTTGG 55
:|||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB
72 AATGGTACGTGATCTTTTGTGAGGTGATGACGTCTTAATGATTTGG 20

RESULT 15
AA951971 584 bp mRNA EST 24-NOV-1998
LD28407.5prime LD Drosophila melanogaster embryo pot2 Drosophila melanogaster cDNA clone LD28407 5prime similar to AF017777: Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (iby), wacław (waw), bobby sox (box), slugish (sig), helicase (hlc), misto (mst), and la costa (lcs) genes, mRNA sequence.
AA951971
NID 93114208
VERSION AA951971.1 GI:3114208
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 584)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brooksstein, P., Lewis, S. and Rubin, G.M.
TITLE
BDF/HMT Drosophila EST Project
JOURNAL
Unpublished (1997)
COMMENT
On Sep 29, 1997 this sequence version replaced g1:1520646.

CONTACT: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
Plate: 284 row: A column: 7
High quality sequence stop: 520.

FEATURES
source
1..584
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1lb="LD Drosophila melanogaster embryo pot2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XLI Blue"
/note="Organ: embryo; Vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pot2."
BASE COUNT
165 a 152 c 180 g 86 t 1 others
ORIGIN

Query Match
Best Local Similarity 52.8%; Score 23.4; DB 40; Length 584;
Matches 28; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY
5 TCGAYTGATHTTTCATACATGATGCGCATCTTTCATATGATTTGGCC 57
:|||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB
514 TCGACTGGCATCTCCACGCTGAGGAGACGCTTCACAGATCTCGGCC 462

Search completed: September 13, 1999, 16:13:49
Job time: 5448 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:19:29 ; Search time 329.34 Seconds

(without alignments)
560.084 Million cell updates/sec

Title: US-09-325-095-1516.SEO

Sequence: 1 AATTGACACACTGGCATGCC.....CGATTTCCTTTTCCTTTTCCTTTT 58

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_cm.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pi1.*
8: gb_pi2.*
9: gb_pi3.*
10: gb_pi4.*
11: gb_pi5.*
12: gb_pi6.*
13: gb_pi7.*
14: gb_pi8.*
15: gb_pi9.*
16: gb_pi10.*
17: gb_pi11.*
18: gb_pi12.*
19: gb_pi13.*
20: gb_pi14.*
21: gb_pi15.*
22: gb_pi16.*
23: gb_pi17.*
24: gb_pi18.*
25: gb_pi19.*
26: gb_pi20.*
27: gb_pi21.*
28: gb_pi22.*
29: gb_pi23.*
30: gb_pi24.*
31: gb_pi25.*
32: gb_pi26.*
33: gb_pi27.*
34: gb_pi28.*
35: gb_pi29.*
36: gb_pi30.*
37: gb_pi31.*
38: gb_pi32.*
39: gb_pi33.*
40: gb_pi34.*
41: gb_pi35.*
42: gb_pi36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	36	62.1	1547	8	GGN132217	Gnetum gn

C	2	35	60.3	638	4	XLU88561	UB8561 Xenopus lae
C	3	35	60.3	35	5	A24382	A24382 d117 adapto
C	4	35	60.3	35	5	A27645	A27645 Hybrid d17
C	5	35	60.3	35	5	A37244	A37244 Sequence 13
C	6	35	60.3	35	5	A40373	A40373 Sequence 29
C	7	35	60.3	35	5	A42335	A42335 Sequence 7
C	8	35	60.3	35	5	A42384	A42384 Sequence 7
C	9	35	60.3	35	5	A46467	A46467 Sequence 5
C	10	35	60.3	35	5	A59198	A59198 Sequence 7
C	11	35	60.3	35	5	A68608	A68608 Sequence 8
C	12	35	60.3	35	5	AR012367	AR012367 Sequence
C	13	35	60.3	35	5	113679	113679 Sequence 35
C	14	35	60.3	35	5	124027	124027 Sequence 9
C	15	35	60.3	35	5	128284	128284 Sequence 13
C	16	35	60.3	35	5	134242	134242 Sequence 19
C	17	35	60.3	35	5	151067	151067 Sequence 25
C	18	35	60.3	35	5	189279	189279 Sequence 12
C	19	35	60.3	723	8	AF062874	AF062874 Arabidops
C	20	35	60.3	646	8	AF062900	AF062900 Arabidops
C	21	35	60.3	9453	9	HS004636	HS004636 Human cyclo
C	22	35	60.3	35	26	E10418	E10418 Primer. 10/
C	23	35	60.3	35	26	E11118	E11118 Oligonucleo
C	24	35	60.3	901	26	E11120	E11120 CDNA 901 bp
C	25	35	60.3	35	26	E11913	E11913 PCR primer
C	26	33	56.9	744	5	A29421	A29421 putative bo
C	27	33	56.9	1193	5	A29423	A29423 putative bo
C	28	33	56.9	1108	5	A29425	A29425 putative bo
C	29	33	56.9	744	5	AR022482	AR022482 Sequence
C	30	33	56.9	1193	5	AR022483	AR022483 Sequence
C	31	33	56.9	1108	5	AR022484	AR022484 Sequence
C	32	33	56.9	744	5	136422	136422 Sequence 13
C	33	33	56.9	1193	5	136423	136423 Sequence 13
C	34	33	56.9	1108	5	136424	136424 Sequence 13
C	35	33	56.9	744	5	188122	188122 Sequence 13
C	36	33	56.9	1193	5	188123	188123 Sequence 13
C	37	33	56.9	1108	5	188124	188124 Sequence 13
C	38	32	55.2	34	5	A59838	A59838 Sequence 8
C	39	32	55.2	3423	11	HS040396	HS040396 Human Stero
C	40	31.8	54.8	35	5	132626	132626 Sequence 11
C	41	31.4	54.1	174311	36	AC004758	AC004758 Drosophila
C	42	30.4	52.4	44219	11	AC005512	AC005512 Homo sapi
C	43	30.2	52.1	226345	42	AC005406	AC005406 , complet
C	44	29	50.0	46275	11	AC003107	AC003107 Human DNA
C	45	28.8	49.7	109339	10	HS137H15	HS137H15 Human DNA

ALIGNMENTS

RESULT 1
GGN132217/c
LOCUS
DEFINITION Gnetum gnetum mRNA for putative MADS domain transcription factor
ACCESSION GCM11.
VERSION AJ132217.1
KEYWORDS gnetum gene; MADS domain transcription factor.
SOURCE Gnetum gnetum.
ORGANISM Gnetum gnetum.
REFERENCE
AUTHORS Winter, K.U., Becker, A., Muenster, T., Kim, J.T., Saedler, H. and Theissen, G.
TITLE MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants
JOURNAL Proc. Natl. Acad. Sci. U.S.A. In press
REFERENCE 2 (bases 1 to 1547)
AUTHORS Winter, K.U., Becker, A., Muenster, T., Kim, J.T., Saedler, H. and Theissen, G.
TITLE MADS-box genes reveal that gnetophytes are more closely related to

confers than to flowering plants

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1547)
AUTHORS Theissen, G.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Theissen G., Molecular Plant Genetics,
Max-Planck-Institut für Breeding Research, Carl-von-Linne Weg 10,
50829 Köln, GERMANY

FEATURES
source
location/Qualifiers
1. 1547
/organism="Gnetum gnemon"
/sub_species="gnemon"
/db_xref="taxon:3382"
159..933
/gene="gsm11"
159..933
/gene="gsm11"
/codon_start=1
/product="putative MADS domain transcription factor GGM11"
/protein_id="CA844457.1"
/db_xref="PID:e1487020"
/db_xref="PID:g5019460"
/db_xref="GI:5019460"
/translation="MGGRVRELKRIKINROYTFKRRNGLLKRAYELSYCLDAEVA
LIIFSRRGLYEGSGAGTLKTEROKCSVALQESNSDROQTMHEHVKIKTKVEI
LORSORHLGEGDEGPISTRELQLEKQVIALQVARKTQAMMDADLKKEKRLQ
EYNSLRKLDEEGVYSNAQIQAPPEPMSNANVIALPLPPQNAVDEPCK
LGIMLLSQAPFGMITRLTYRGSNSNIPANNQ"

BASE COUNT 465 a 308 c 372 g 402 t

ORIGIN

Query Match 62.1%; Score 36; DB 8; Length 1547;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 TGACTGAGTCGACATCGATTGTTTTTTTTTTT 58
|||||

Db 1215 TGACTGAGTCGACATCGATTGTTTTTTTTTTT 1180

RESULT 2
XLU88561 638 bp mRNA VRT 16-MAY-1997
LOCUS Xenopus laevis E2 ubiquitin conjugating enzyme (Ubc9) mRNA,
DEFINITION complete cds.
ACCESSION U88561
NID 52078330
VERSION U88561.1 GI:2078330
KEYWORDS African clawed frog.
SOURCE Xenopus laevis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
Xenopus.

REFERENCE 1 (bases 1 to 638)
AUTHORS Saitoh, H., Pu, R., Cavenagh, M. and Dasco, M.
TITLE RanBP2 associates with Ubc9p and a modified form of RanGAP1
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (8), 3736-3741 (1997)
MEDLINE 97268639
REFERENCE 2 (bases 1 to 638)
AUTHORS Saitoh, H., Pu, R., Cavenagh, M. and Dasco, M.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-1997) LME/NICHD, NIH, Bldg. 18, Rm. 101, Bethesda,
MD 20892, USA

FEATURES
source
location/Qualifiers
1. 638
/organism="Xenopus laevis"
/db_xref="taxon:8335"
/cell_type="oocyte"
107..583
/gene="Ubc9"
107..583

/gene="ubc9"
/note="Ubc9p"
/codon_start=1
/product="E2 ubiquitin conjugating enzyme"
/protein_id="AAB57736.1"
/db_xref="PID:g2078331"
/db_xref="GI:2078331"
/translation="MSGIALSLAERKARKRDPFGVAVPTKPNPQGTMINMNECA
IPKRGTEWEGSEFLRLMFLKDDPSSPKCKFEPLFLHPNVYPSGIVCLSLDEKDK
MRPATIKQILGILQIELNENPDIQDPAQAEAVTYICNRRVEKRRVQAKKFAPS"

BASE COUNT 214 a 131 c 147 g 146 t

ORIGIN

Query Match 60.3%; Score 35; DB 4; Length 638;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 58
|||||

Db 638 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 604

RESULT 3
A24382 35 bp DNA PAT 21-MAR-1995
LOCUS dlt1 adaptor primer.
DEFINITION A24382
ACCESSION 9904477
NID 9904477
VERSION A24382.1 GI:904477
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 35)
AUTHORS Samlentos, P., De Taxis du Poet, P., Nitli, G. and Socheri, E.
TITLE Anti-thrombin polypeptides
JOURNAL Patent: EP 0501821 A 22 02-SEP-1992;
FARMITALIA CARLO ERBA S.r.L.

FEATURES
source
location/Qualifiers
1. 35
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 5 a 5 c 5 g 20 t

ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 58
|||||

Db 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 4
A27645 35 bp DNA PAT 04-JUN-1995
LOCUS Hybrid dlt1-adaptor primer.
DEFINITION A27645
ACCESSION 91248481
NID 91248481
VERSION A27645.1 GI:1248481
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 35)
AUTHORS POLYPHENOL OXIDASE GENES
TITLE Patent: WO 9302195-A 1 04-FEB-1993;
JOURNAL Location/Qualifiers
FEATURES
source
1. 35

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN
/organism="synthetic construct"
/db_xref="taxon:32630"

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 58
|||||
Db 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 5
LOCUS A37244 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 13 from Patent WO9403609.
ACCESSION A37244
NID 92294355
VERSION A37244.1 GI:2294355
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 35)
AUTHORS Parker, P. J., Goode, N. T., Nurse, P. M. and Waterfield, M. D.
TITLE EDUCAROTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
*OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 13 17-FEB-1994;
TMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
source 1..35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN
/organism="unidentified"
/db_xref="taxon:32644"

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 58
|||||
Db 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 6
LOCUS A40373 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 29 from Patent WO9425489.
ACCESSION A40373
NID 92296422
VERSION A40373.1 GI:2296422
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 35)
AUTHORS Mohapatra, S. S. and Sehon, A. H.
TITLE DOWNREGULATION OF ALLERGEN-SPECIFIC IMMUNE RESPONSE
JOURNAL Patent: WO 9425489-A 29 10-NOV-1994;
UNIV MANITOBA (CA)
COMMENT Other publication AU 6674094 941121.
FEATURES
source 1..35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN
/organism="unidentified"
/db_xref="taxon:32644"

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 58
|||||
Db 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 7
LOCUS A42335 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 7 from Patent WO9502057.
ACCESSION A42335
NID 92297812
VERSION A42335.1 GI:2297812
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 35)
AUTHORS Gusterson, B. A., Crompton, M. R., Mitchell, P. J., Barker, K. T.,
Kamaleri, T., Page, M. J. and Spence, P.
TITLE PROTEIN TYROSINE KINASE AND LIGANDS THEREOF
JOURNAL Patent: WO 9502057-A 7 19-JAN-1995;
CANCER RES INST (GB)
COMMENT Other publication AU 7080994 950206.
FEATURES
source 1..35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN
/organism="unidentified"
/db_xref="taxon:32644"

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 58
|||||
Db 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 8
LOCUS A42384 35 bp DNA PAT 06-MAR-1997
DEFINITION Sequence 7 from Patent WO9502187.
ACCESSION A42384
NID 92297858
VERSION A42384.1 GI:2297858
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 35)
AUTHORS Gusterson, B. A., Crompton, M. R., Mitchell, P. J., Barker, K. T.,
Martindale, J. E., Page, M. J. and Spence, P.
TITLE CELL GROWTH FACTOR RECEPTORS
JOURNAL Patent: WO 9502187-A 7 19-JAN-1995;
CANCER RES INST (GB)
COMMENT Other publication AU 7081094 950206.
FEATURES
source 1..35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN
/organism="unidentified"
/db_xref="taxon:32644"

Query Match 60.3%; Score 35; DB 5; Length 35;

Training Match	60.38;	Score 35;	DB 5;	Length 40;
Best Local Similarity	100.00;	Pred. No.	0.015;	

Db 1 GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT 35

RESULT 13
LOCUS 113679 35 bp DNA PAT 08-AUG-1995
DEFINITION Sequence 35 from patent US 5439820.
ACCESSION 113679
MID 9996745
VERSION 113679.1 GI:996745
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Samientos,P., De Taxis du Poet,P., Nitli,G. and Scacheri,E.
TITLE Anti-thrombin polypeptides
JOURNAL Patent: US 5439820-A 35 08-AUG-1995;
FEATURES Location/Qualifiers
source 1..35
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred.No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

RESULT 14
LOCUS 124027 35 bp DNA PAT 21-NOV-1996
DEFINITION Sequence 9 from patent US 5541110.
ACCESSION 124027
MID 91603897
VERSION 124027.1 GI:1603897
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Siegall,C.B.
TITLE Cloning and expression of a gene encoding bryodin 1 from Bryonia dioica
JOURNAL Patent: US 5541110-A 9 30-JUL-1996;
FEATURES Location/Qualifiers
source 1..35
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred.No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

RESULT 15
LOCUS 128284 35 bp DNA PAT 30-OCT-1996
DEFINITION Sequence 13 from patent US 5569830.
ACCESSION 128284
MID 91819060
VERSION 128284.1 GI:1819060
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

Unclassified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
TITLE Plant inhibitors of fungal polygalacturonases and their use to control fungal disease
JOURNAL Patent: US 5569830-A 13 29-OCT-1996;
FEATURES Location/Qualifiers
source 1..35
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred.No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

Search completed: September 13, 1999, 16:19:30
Job time: 4482 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:21:48 ; Search time 133.08 Seconds
(without alignments)
109.041 Million cell updates/sec

Title: US-09-325-095-1516.SEQ

Sequence: 1 AATTCACACACTGCGATGCC.....CGATTTTTCCTTTTTCCTTTT 58

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	60.3	35	1	036668	PPO adaptor primer
2	35	60.3	35	1	052503	Helminth aminopept
3	35	60.3	35	1	059015	RACE PCR adaptor-d
4	35	60.3	35	1	057021	Ptdins 3-kinase ad
5	35	60.3	35	1	071021	Primer for amplify
6	35	60.3	35	1	065873	T cell protein CD4
7	35	60.3	35	1	065594	HER4 primer XSC717
8	35	60.3	35	1	078807	TCR alpha-chain dt
9	35	60.3	35	1	084786	PKR2 reverse-tran
10	35	60.3	35	1	090167	Pollen allergen ge
11	35	60.3	35	1	089878	PCR primer for clo
12	35	60.3	35	1	705321	Oligonucleotide pr
13	35	60.3	35	1	094244	Cysteine proteinas
14	35	60.3	35	1	110696	Trypsin-like enzym
15	35	60.3	35	1	110698	Trypsin-like enzym
16	35	60.3	35	1	116211	Primer XSC-117. O1
17	35	60.3	35	1	710276	Human EPII/placent
18	35	60.3	35	1	703701	Oligo-dt primer fo
19	35	60.3	35	1	711753	Primer for human k
20	35	60.3	35	1	074002	Oligo dt-adaptor p
21	35	60.3	35	1	735802	Coffee bean alpha-
22	35	60.3	35	1	728797	MMV reverse trans
23	35	60.3	35	1	742690	Primer #1 for huma
24	35	60.3	35	1	742699	Primer #6 for Vers
25	35	60.3	35	1	743345	Primer #1 for acyl
26	35	60.3	35	1	749441	Pear polygalacturo
27	35	60.3	35	1	736629	Primer for human p
28	35	60.3	35	1	761267	NAG tumour rejecti
29	35	60.3	35	1	747688	Adaptor primer, B2
30	35	60.3	35	1	772469	dh17 anchor primer
31	35	60.3	35	1	776809	Primer EAL3 for le
32	35	60.3	35	1	773128	Primer for mouse T
33	35	60.3	35	1	799144	Primer for S-adeno
34	35	60.3	35	1	788344	Primer EAL3 for As
35	35	60.3	35	1	V09259	Clavin cDNA PCR pr
36	35	60.3	35	1	V25554	Vascular endotheli
37	35	60.3	35	1	V25542	Vascular endotheli
38	35	60.3	35	1	V28092	Human galectin-4 r
39	35	60.3	35	1	V53244	Ecdysis triggering
40	35	60.3	35	1	V56016	D. discoideum DP1
41	35	60.3	35	1	V74046	Adaptor-oligo(dt)
42	35	60.3	35	1	V64425	Mouse Praja-1 gene
43	35	60.3	35	1	X24713	Adaptor primer use

ALIGNMENTS

C	44	33	56.9	744	1	030669
C	45	33	56.9	1193	1	030670
ALIGNMENTS						
RESULT	1					
ID	036668					
AC	036668					
DT	09-JUN-1993					
DE	PPO adaptor primer #1.					
KM	Polyphenol oxidase; PPO; catalyzt; browning; fruit; plastid; vacuole;					
KW	transfom; coffee; tea; black olives; grapevine; chloroplast; apple;					
KW	transit peptide; recombinant plasmid; PCR; primer; amplify; broad bean;					
KW	potato; polymerase chain reaction; ss.					
OS	Synthetic.					
PN	W0302195-A.					
PD	04-FEB-1993.					
PF	16-JUL-1992; A00356.					
PR	17-JUL-1991; AU-007248.					
PA	(CSIR) COMMONWEALTH SCI & IND RES ORG.					
PI	Dry IB, Robinson SP;					
DR	WPI: 93-058792/07.					
FT	DNA encoding polyphenol oxidase polypeptide or fragment - useful					
PT	for modifying the oxidase activity in fruit and vegetables to					
PS	decrease or enhance browning					
CS	Claim 15: Page 23; 44pp; English.					
CC	The sequences given in 036668-69 are adaptors which were used in the					
CC	cloning of the polyphenol oxidase (PPO) enzyme genes from various					
CC	plants. The PPO genes were isolated, and recombinant plasmids for					
CC	transformation of plant cells were produced by PCR using the primers					
CC	given in 036678-78. PPO is thought to be the predominant catalyst in					
CC	browning of fruit caused by injury or damage. PPO is localised in the					
CC	plastids of plant cells whereas the phenolic substrates of the enzyme					
CC	are stored in the plant cell vacuole. This compartmentation prevents					
CC	the browning reaction from occurring unless the plant cells are damaged					
CC	and the enzyme and the substrate are mixed. The PPO gene sequences					
CC	could be used to construct synthetic genes which may be used to					
CC	transform plants to decrease expression of the enzyme gene. In some					
CC	instances, eg. coffee, tea, black olives etc., it is desirable to					
CC	increase the level of PPO to produce desired levels of browning or					
CC	changes in flavour compounds. The grapevine PPO gene codes for an					
CC	additional 103 amino acids upstream of the N-terminus of the mature					
CC	protein. This region has the properties of a chloroplast transit					
CC	peptide and is most likely responsible for targeting of the protein					
CC	to be imported into the chloroplast and processed to produce mature					
CC	PPO. Transformation of plants with this gene may therefore result					
CC	in correct targeting and maturation of the grapevine PPO in other					
CC	species and result in accumulation of active grapevine PPO enzyme in					
CC	the plastids of these tissues.					
Sequence	35 BP;	5 A;	5 C;	5 G;	20 T;	
Query Match						
Best local similarity 60.3%; Score 35; DB 1; Length 35;						
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	24	GACCTGAGTCGACATGATTTTTCCTTTTTCCTTTT 58				
DB	1	GACTCGAGTCGACATGATTTTTCCTTTTTCCTTTT 35				
RESULT 2						
ID	052503					
AC	052503					
DT	31-MAY-1994					
DE	Helminth aminopeptidase H11-1 adaptor primer.					
KW	Helminth aminopeptidase; H11-3; H11-2; H11-1; Integral membrane;					
KW	alpha-amino acyl peptide hydrolase (microsome); multigene family;					
KW	antigen; vaccine; parasite; human; virus; microbe; PCR; primer;					

KW polymerase chain reaction; amplify; Haemonchus contortus; ss.
 OS Synthetic.
 PN WO9323542-A.
 PD 25-NOV-1993.
 PR 07-MAY-1993: G00943.
 PF 08-MAY-1992: GB-009993.
 PA (AGRI-) AGRIC & FOOD RES COUNCIL.
 PI Graham M, Knox DP, Munn EA, Newton SE, Oliver JJ,
 PI Smith TS;
 DR WPI: 93-386574/48.
 PT New DNA encoding amino-peptidase from *Haemonchus contortus* and
 PT derived peptide(s) - useful in protective vaccines, for new
 PT vectors, transformed cells and oligosaccharide, for incorporation
 PT in virus or microbe
 PS Example: Page 32: 137pp; English.
 SS The sequences given in 052503-11 are primers which were used to
 SS amplify cDNA clones of the helminth aminopeptidase genes H11-1, -2
 CC and -3. The amplified sequences encode H11D which is a protein
 CC doublet which shows homology to a family of integral membrane amino-
 CC peptidases. The differences between the amplified clones can be
 CC attributed to different mRNAs of the multigene family, and also to
 CC different variants of the H11D-encoding sequence being present at
 CC different stages of the life cycle, or in strains differing in
 CC geographical origin. Antigenic fragments of the aminopeptidases
 CC encoded by the H11 genes may be used in vaccines to stimulate immune
 CC response against helminth parasites in humans or other animals. These
 CC DNA sequences may be incorporated into a virus or microbe and used in
 CC a similar manner.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 24 GACTCGAGTCGACATCGATTTT TTTT TTTT 58
 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 3
 OS9015
 ID 059015 standard; DNA: 35 BP.
 AC 059015.
 DT 12-APR-1994 (first entry)
 DE RACE PCR adaptor-dt.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ss; amplification; p110.
 OS Synthetic.
 PN WO9321328-A.
 PD 28-OCT-1993.
 PR 13-APR-1993: G00761.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G,
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1; Page 39; 146pp; English.
 SS An SGBAF-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with pSV3neo. Total RNA was
 CC isolated from the SGBAF-1 and a cDNA library constructed. RACE PCR
 CC was performed by synthesizing a first strand cDNA from random hexamers
 CC on SGBAF-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-
 CC dt as primers. Products were fractionated using an agarose gel. The
 CC DNA was isolated from the gel and subjected to PCR using oligo 2280 and
 CC adaptor as primers. A product of 350 bp was further sequenced.
 CC See also 05115-6, 059012-23 and 057522-3.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 24 GACTCGAGTCGACATCGATTTT TTTT TTTT 58
 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 4
 OS7021
 ID 057021 standard; DNA: 35 BP.
 AC 057021.
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase adaptor dt primer.
 KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KW ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Synthetic.
 PN WO9403609-A.
 PD 17-FEB-1994.
 PR 05-AUG-1993: G01651.
 PA (IMCR-) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Example 1; Page 15; 71pp; English.
 SS The sequences given in 057020-23 are primers which were used in the
 CC isolation of the phosphatidylinositol (Ptdins) 3-kinase cDNA. The
 CC amplified sequence was placed under the regulatory control of the nmt
 CC promoter. These sequences were transformed into *Schizosaccharomyces*
 CC pombe cells in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the ptdins
 CC plasmids grow as the parental strain. In the absence of thiamine the
 CC nmt promoter functions and the ptdins is induced. Ptdins activity is
 CC substantially increased under these conditions. Cells containing
 CC constructs such as this, are useful in assays for detecting compounds
 CC involved in cell growth regulation. It is also used as the basis for
 CC detecting compounds for treating cancers and the formation of blood
 CC vessel plaques.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 24 GACTCGAGTCGACATCGATTTT TTTT TTTT 58
 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 5
 OS71021
 ID 071021 standard; DNA: 35 BP.
 AC 071021.
 DT 26-SEP-1994 (first entry)
 DE Primer for amplifying peach polygalacturonase coding sequence.
 KW Ripening; melting; pectins; breeding; Prunus; Rosaceae;
 KW endopolygalacturonase; EndoPG; polygalacturonase; PG; ss.
 OS Synthetic.
 PN WO9405795-A.
 PD 17-MAR-1994.
 PR 27-AUG-1993: AU0442.
 PA (CSIR-) COMMONWEALTH SCI & IND RES ORG.
 PA (PACI-) PACIFIC SEEDS PVT LTD.
 PI Brady CJ, Lee E, Lester DR, Orr G, Speirs J;

DR WPI: 94-101199/12.
PT DNA encoding peach polygalacturonase (Pg) and production of
PT recombinant peach Pg in rosaceous plants - for production of
PT plants with 'melting' or 'non-melting' phenotype
PS Disclosure: Page 15; 51pp; English.
CC During early ripening of peaches, tissue firmness decreases
CC slowly and progressively. Towards the end of ripening, loss of
CC tissue firmness is rapid. This second stage of softening is called
CC the 'melting' stage. Fruit of peach varieties used for canning do
CC not have a 'melting' phase of softening. Ripen fruit remain
CC relatively firm and maintain their shape throughout processing.
CC Fruit of 'melting' varieties show an increase in activity of
CC endopolygalacturonase (EndoPG) during ripening. Polygalacturonase
CC (PG) is believed to contribute to fruit softening through its action
CC on intercellular and cell wall pectins. This primer was used to
CC synthesise a first strand cDNA molecule for further amplification
CC by PCR. See Q56765-67.
SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTGTGTTT 58
|||||
DB 1 GACTCGAGTCGACATCGATTTTGTGTTT 35

RESULT 6
O63873
ID Q63873 standard; cDNA; 35 BP.
AC Q63873;
DT 04-NOV-1994 (first entry)
DE T cell protein CD4 adaptor-drl1 primer.
KW Cat; feline immunodeficiency virus; CD4 positive T lymphocyte;
KW cellular receptor; T cell protein; polymerase chain reaction; PCR; ss.
OS Synthetic.
PN FR2696188-A.
PD 01-APR-1994.
PF 25-SEP-1992: 011496.
PR 25-SEP-1992: FR-011496.
PA (UYPA-) UNIV CURIE PARIS VI P & M.
PI DE Parseval A, Klatzmann D, Salmon P, Sonigo P;
DR WPI: 94-128285/16.
PT Feline CD4 protein and corresp. DNA - for study of feline
PT immunodeficiency virus and for therapy
PS Disclosure: Page 5; 25pp; French.
CC The primers Q63871 and Q63872 were used to obtain cDNA coding for
CC feline CD4 protein by polymerase chain reaction. For rapid
CC amplification of the ends of the cDNA, a hybrid adaptor-oligo(dT)
CC primer was also used. Cat lymph node cDNA was amplified and the
CC full-length CD4 coding sequence Q63870 was obtained from several
CC partial clones.
SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTGTGTTT 58
|||||
DB 1 GACTCGAGTCGACATCGATTTTGTGTTT 35

RESULT 7
O63594
ID Q63594 standard; DNA; 35 BP.
AC Q63594;
DT 12-JAN-1995 (first entry)
DE HER4 primer XScrl7.
KW Human epidermal growth factor receptor; HER; EGF; tyrosine kinase;

KW cancer; neuronal tissue; muscle tissue; neoplasm;
KW carcinoma; primer; probe; PCR; ss.
OS Synthetic.
PN EP-599274-A.
PD 01-JUN-1994.
PR 23-NOV-1993; 118837.
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Culouscou J, Ploman GD, Shoyab M;
DR WPI: 94-169599/21.
PT New recombinant nucleic acid expressing HER4 - a new receptor
PT tyrosine kinase expressed in some cancer cells, and related
PT vectors, antibodies, ligands etc, for diagnosis and treatment of
PT cancers
PS Example: Page 16; 104pp; English.
CC The primers given in Q63594-6 were used in the semi-quantitative
CC PCR detection of HER4.
CC HER4 is the fourth member of the EGF-family of tyrosine kinases and
CC is expressed in some human cancers and in some tissues of neuronal
CC or muscle origin. HER4 polynucleotides, opt. labelled, are useful
CC in assays (e.g. of HER4 mRNA to detect certain neoplasms, esp.
CC breast carcinoma) and as primers in PCR or as probes.
SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTGTGTTT 58
|||||
DB 1 GACTCGAGTCGACATCGATTTTGTGTTT 35

RESULT 8
O78807
ID Q78807 standard; cDNA; 35 BP.
AC Q78807;
DT 12-JUN-1995 (first entry)
DE TCR alpha-chain dT17 adaptor.
KW Allergen; immunotherapy; T-cell receptor; TCR; CDR3;
KW complementarity determining region-3; immunosuppressor;
KW antibody engineering; suppressor T-lymphocyte; Ts; PCR;
KW polymerase chain reaction; amplification; adaptor; ss.
OS Synthetic.
PN WO9425489-A.
PD 10-NOV-1994.
PF 26-APR-1994; CA0228.
PR 26-APR-1993; GB-008581.
PA (UYMA-) UNIV MANITOBA.
PI Mohapatra SS, Sehon AH;
DR WPI: 94-358193/44.
PT Synthetic peptide(s) for effecting immuno-therapy - have an amino
PT acid sequence comprising a portion of the CDR3 region of a T-cell
PT receptor
PS Disclosure: Page 43; 78pp; English.
CC Antibody production against allergenic antigen (Ag) is specifically
CC suppressed by treatment with the junctional segment of the alpha
CC and/or beta chain of the Ag receptor of suppressor T (Ts) cells
CC (Tcr) induced by tolerogenic Ag-Peg conjugates. cDNAs encoding
CC the alpha and beta chains of TCRs of cloned Ts cells specific for
CC ovalbumin and human monoclonal (myeloma) IgG were produced by PCR
CC using the primers given in Q78798-805 and the adaptors given in
CC Q78806-07.
SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTGTGTTT 58
|||||

```

Db 1 GACTCGAGTCGACATCGATTTTTCCTTTTTCCTTTT 35
RESULT 9
ID 084786
AC 084786 standard; DNA: 35 BP.
DE 17-AUG-1995 (first entry)
PTK22 reverse-transcription primer.
KW Protein-tyrosine kinase; PK; discoidin domain receptor; cancer;
breast tumor; mammary carcinoma; diagnosis; prognosis; therapy; ss.
OS Synthetic.
PN WO9502187-A.
PD 19-JAN-1995.
PE 08-JUL-1994; GB-014271.
PR 09-JUL-1993; GB-014271.
(PATENT) CANCER RES INST.
PA (WELLS ) WELLCOME FOUND LTD.
PI Baker KT, Crompton MR, Gusterson BA, Martindale JE;
PI Mitchell PJ, Page MJ, Spence P;
PI MPI: 95-066991/09.
PT Method for screening substances, using protein tyrosine kinase -
PT for potential utility as therapeutic agents for cancer
PS Disclosure: Page 34; 51pp; English.
CC cDNA derived from tumor metastatic tissue was amplified using
CC primers (given in 084783-84) based on sequences (R7101, R7103)
CC associated with protein-tyrosine-kinases (PTK). Novel PTK22 was
CC identified in an isolated subclone. The 3' sequence of PTK22 was
CC obtained by reverse transcription (using the primer of 084786) and
CC PCR amplification (primers 084787-88) of RNA of human breast
CC carcinoma cell line MDA MG 468. The partial DNA sequence of PTK22
CC is given in 084782.
SO Sequence 35 BP: 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTTCCTTTTTCCTTTT 58
Db 1 GACTCGAGTCGACATCGATTTTTCCTTTTTCCTTTT 35
RESULT 10
ID 090167
AC 090167 standard; DNA: 35 BP.
DE 01-NOV-1995 (first entry)
PT Pollen allergen gene primer 8.
KW Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;
KW desensitizer; Cryptomeria japonica; polymerase chain reaction;
KW PCR; primer; ss.
OS Synthetic.
PN EP-655500-A.
PD 31-MAY-1995.
PE 03-NOV-1994; 308117.
PR 05-NOV-1993; JP-299151.
PR 20-DEC-1993; JP-344596.
PR 27-DEC-1993; JP-346814.
(PATENT) HAYASHIBARA SEIBUTSU KAGAKU.
PI Kurimoto M, Namba M, Torigoe K;
PI MPI: 95-195588/26.
PT New Japanese cedar pollen allergen polypeptide - and DNA coding for it.
PT useful for treatment and diagnosis of cedar pollen allergy
PS Disclosure: Page 11; 41pp; English.
CC The primers given in 090166-69, based on a portion of cDNA clone
CC SC09 or on previously isolated sequences, were used for the PCR
CC amplification of cDNA derived from Japanese cedar pollen, to obtain
CC clone SC50 bearing a partial sequence (nt 1-198 of the sequence
CC given in 090156) of the pollen allergen gene.
SO Sequence 35 BP: 5 A; 5 C; 5 G; 20 T;

```

```

Query Match          60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches    35; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

QY      24 GACTCGAGTCGACATCATGATTCTTTTCTTTTTTTT 58
        |||
DB       1 GACTCGAGTCGACATCATGATTCTTTTCTTTTTTTT 35

RESULT  11
ID      C089878 standard; DNA; 35 BP.
AC      C089878;
DT      31-JAN-1997 (Revised)
DT      25-OCT-1995 (first entry)
RW      PCR primer for cloning bryodin-2.
KW      Ribosome inactivating protein; cell death; proliferation; tumour;
KW      cancer; virus; malaria; trypanosomiasis; inflammation; autoimmune;
OS      disease; purging; bone marrow; ss.
OS      Synthetic.
PN      W09511977-A.
PD      04-MAY-1995.
PF      25-OCT-1994; U12382.
PR      25-OCT-1993; US-141891.
PR      20-OCT-1994; US-324301.
PA      (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI      Gavlak SL, Marguardt H, Siegall CB;
PI      WP1: 95-178870/23.
PT      New ribosome inactivating protein bryodin-2 and related
PT      conjugates - for killing target cells, e.g. in treatment of
PT      cancer, also related nucleic acid, vectors and transformed cells
PS      Disclosure: Page 43; 81pp; English.
CC      The sequence is that of a PCR primer used to isolate a ribosome
CC      inactivating protein termed bryodin-2, from Bryonia dioica. Conjugates
CC      of the protein with a ligand may be used to kill target cells (i.e. those
CC      to which the ligand component binds specifically) or to inhibit
CC      proliferation of tumour cells. Typical applications include the
CC      treatment of cancer, viral infections, malaria, trypanosomiasis,
CC      inflammatory or autoimmune diseases, including in vitro purging of e.g.
CC      bone marrow.
CC      See also Q89874-B6.
CC      Updated 31-JUN-1997 due to incorrect cross reference numbers.
SQ      Sequence    35 BP:   5 A;     5 C;     5 G;    20 T;

Query Match          60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches    35; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

QY      24 GACTCGAGTCGACATCATGATTCTTTTCTTTTTTTT 58
        |||
DB       1 GACTCGAGTCGACATCATGATTCTTTTCTTTTTTTT 35

RESULT  12
ID      T05321 standard; DNA; 35 BP.
AC      T05321;
DT      13-APR-1996 (first entry)
DT      Oligonucleotide probe for human prostacyclin-synthase.
RW      DNA probe; prostacyclin-synthase; PCR; polymerase chain reaction;
KW      DNA primer; prostaglandin I2; circulatory disease; therapeutic;
KW      diagnosis; gene therapy; ss.
OS      Synthetic.
PN      W09530013-A1.
PD      09-NOV-1995.
PF      27-APR-1995; J00838.
PR      28-APR-1994; JP-114316.
PA      (TANABE/) TANABE T.
PI      Tanabe T.;
PI      WP1: 95-393084/50.
PT      Human prostacyclin synthase and DNA encoding it - useful in the

```

PT Investigation and treatment of diseases characterised by reduced prostaglandin 12 production.

PS Disclosure: Page 34; 71pp; Japanese.

CC DNA primers (T05317-20; T05322; T05326-27) are used to screen human genomic lung cell line W138 and human arterial endothelial cell CC cDNA libraries for the isolation of a prostacyclin synthase (PGIS) CC coding sequence (see T05316). DNA probes (T05321 and T05323) are used in the construction of plasmid pHPGIS1, encoding the complete PGIS CC sequence. This plasmid was used to transfect human 293 cells for CC PGIS peptide expression. DNA encoding human PGIS, vectors containing CC it, and PGIS itself, may be administered to patients to increase CC prostaglandin 12 (PGI2) production to treat diseases characterized CC by reduced PGI2 levels or by an imbalance between PGI2 and CC thromboxane A2 levels, such as circulatory diseases (thrombosis, CC angina pectoris, arteriosclerosis, myocardial infarction). The CC DNA and protein are also useful in disease diagnosis.

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTTCCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTTTCCTTTT 35

RESULT 13

ID 094244 standard; DNA; 35 BP.

AC 094244;

DT 14-MAY-1996 (first entry)

KW Cysteine proteinase active site antisense polyT primer.

KW Primer: amplify; polymerase chain reaction; PCR; Haemonchus contortus;
KM cysteine proteinase; DM.2; DM.3; DM.4; human; DM.4a; DM.5; antigen;
KW vaccine; helminth; parasite; DM.1; ruminant; gastro-intestinal tract; ss.

OS Synthetic.

PN W09526402-A1.

PD 05-OCT-1995.

PF 24-MAR-1995; G00065.

PR 25-MAR-1994; GB-005990.

PA (MLC) MALINKRODT VETERINARY INC.

PI Knox DP, Murray J, Redmond D, Smith SK, Smith WD;

DR WPI; 95-351322/45.

PT Protective helminth parasite antigen - used in vaccine directed PT against parasitic nematodes of mammalian gastro-intestinal tract

PS Example 16; Fig 15; 79pp; English.

CC The sequences given in 094240-45 are primers which are based on CC previously published sequences derived from the canonical Haemonchus CC contortus cysteine proteinase molecule. These primers were used in CC the cloning of cDNA fragments from the cysteine proteinase gene, CC such as DM.1, DM.2, DM.3, DM.4, DM.4a and DM.5 (see also 094246-51).

CC The amplified fragments may be expressed in a recombinant cell for CC the production of antigens. These antigens may be used in the CC preparation of a vaccine against helminth parasites in a human or CC non-human animal.

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTTCCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTTTCCTTTT 35

RESULT 14

T10696

ID T10696 standard; DNA; 35 BP.

AC T10696;

DT 26-MAY-1996 (first entry)

DE Trypsin-like enzyme oligonucleotide TRX-10.

KW Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;
KW asthma; VIP; vasoactive intestinal peptide; influenza virus;
KW protease; primer; PCR; amplification; ss.

OS Synthetic.

PN A09527248-A.

PD 08-FEB-1996.

PF 31-JUL-1995; 027248.

PR 29-JUL-1994; JP-178607.

PA (TEIJU) TEIJUN LTD.

PI Masuda K, Ogawa H, Suga T, Sugimoto Y, Takagi K;

PI Yamaoka K, Yasuoka S;

DR WPI; 96-117356/13.

PT Nucleic acid sequence encoding trypsin-like enzyme - which digests PT fibrinogen, used as expectorant in treatment of respiratory

PT diseases, e.g. bronchial asthma

PS Example 9, Page 44; 65pp; English.

CC Example 9 describes the cloning of cDNA region encoding CC trypsin-like enzyme isolated from cough phlegm.

CC Four oligonucleotides are used for PCR: TRX-1 (T10694), TRX-8 CC (T10695), TRX-10 (T10696) and TRX-11 (T10697).

CC TRX-1 corresponds to from A1 to A23 of the sequence given in CC T10693, which is part of a cDNA encoding the trypsin-like CC enzyme. TRX-8 corresponds to from G16 to T40 of T10693.

CC TRX-10 is capable of annealing to the 3'-terminus of poly(A)+ RNA. CC After amplification, plasmid p19-33 was obtained. p19-33 encodes CC part of the N-terminus amino acid sequence 20 residues of the CC trypsin-like enzyme isolated from the cough phlegm (see T10698).

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTTCCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTTTCCTTTT 35

RESULT 15

ID T10698 standard; cDNA to mRNA; 901 BP.

AC T10698;

DT 26-MAY-1996 (first entry)

DE Trypsin-like enzyme N-terminal DNA fragment from p19-33.

KW Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;
KW asthma; VIP; vasoactive intestinal peptide; influenza virus;
KW protease; primer; PCR; amplification; ds.

OS Homo sapiens.

PN A09527248-A.

PD 08-FEB-1996.

PF 31-JUL-1995; 027248.

PR 29-JUL-1994; JP-178607.

PA (TEIJU) TEIJUN LTD.

PI Masuda K, Ogawa H, Suga T, Sugimoto Y, Takagi K;

PI Yamaoka K, Yasuoka S;

DR WPI; 96-117356/13.

PT Nucleic acid sequence encoding trypsin-like enzyme - which digests PT fibrinogen, used as expectorant in treatment of respiratory

PT diseases, e.g. bronchial asthma

PS Example 9, Page 44-45; 65pp; English.

CC Example 9 describes the cloning of cDNA region encoding CC trypsin-like enzyme isolated from cough phlegm.

CC Four oligonucleotides are used for PCR: TRX-1 (T10694), TRX-8 CC (T10695), TRX-10 (T10696) and TRX-11 (T10697).

CC TRX-1 corresponds to from A1 to A23 of the sequence given in CC T10693, which is part of a cDNA encoding the trypsin-like CC enzyme. TRX-8 corresponds to from G16 to T40 of T10693.

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:13:49 ; Search time 849.94 Seconds

(Without alignments)
134,606 Million cell updates/sec

Title: US-09-325-095-1516.SEQ

Perfect score: 58
Sequence: 1 AATTCACACACTGGCAGCC.....CGATTTTTCCTTTTTCCTTTT

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	38	65.5	337	32	AA340759	AA340759 EST146034
2	28.2	48.6	241	44	A1270969	A1270969 gw52f12.x
3	27.8	47.9	541	47	A1479049	A1479049 tm30e03.x
4	27.8	47.9	474	48	A1568396	A1568396 th33g06.x
5	27	46.6	477	43	A1229956	A1229956 EST122651
6	26.8	46.2	397	25	N89796	N89796 db34b04..s1
7	26.8	46.2	498	39	AA833826	AA833826 od62c05.s
8	26.8	46.2	425	39	AA902569	AA902569 o35d09.s
9	26.8	46.2	765	41	A1026930	A1026930 ov82e09.x
10	26.8	46.2	456	42	A1079505	A1079505 oy95f01.x
11	26.8	46.2	511	44	A1242489	A1242489 q116f01.x
12	26.8	46.2	723	45	A1378815	A1378815 lc67d11.x
13	26.8	46.2	527	46	A1445584	A1445584 t326a11.x
14	26.8	46.2	600	48	A1568176	A1568176 tn66d10.x
15	26.6	45.9	207	50	AV025932	AV025932 AV025932
16	26.4	45.5	452	25	N50597	N50597 y789e06.r1
17	26.4	45.5	435	28	C18175	C18175 C18175 Huma
18	26.4	45.5	399	34	AA509647	AA509647 vhs1g12.r
19	26.4	45.5	499	37	AA675277	AA675277 v705f10.y
20	26.4	45.5	390	44	A1255155	A1255155 u132a11.y
21	26.4	45.5	616	51	A1725226	A1725226 1245 PRIF
22	26	44.8	478	26	W17126	W17126 zbl7f05.r1
23	26	44.8	492	28	AA114945	AA114945 z88a08.r
24	26	44.8	338	29	AA164347	AA164347 zp04g09.r
25	26	44.8	180	36	C61881	C61881 C61881 Yuj1
26	26	44.8	352	49	A1640742	A1640742 wa15a02.x
27	25.8	44.5	698	26	W20094	W20094 db39d09..r1
28	25.8	44.5	570	29	AA145612	AA145612 ms07h10.r
29	25.8	44.5	357	37	AA668836	AA668836 ab91g02.s
30	25.8	44.5	656	44	A1322793	A1322793 ms07h10.y
31	25.6	44.1	678	44	A1295887	A1295887 LP09649.5
32	25.6	44.1	287	48	A1545931	A1545931 fb66b05.y
33	25.6	44.1	534	48	A1557358	A1557358 PT2.1.6.B
34	25.6	44.1	1100	51	AU050243	AU050243 AU050243
35	25.4	43.8	317	41	A1002468	A1002468 oq83h06.s
36	25.4	43.8	349	44	A1305227	A1305227 qm13f04.x
37	25.4	43.8	111	44	A1313326	A1313326 ca78e03.x
38	25.4	43.8	380	45	A1348883	A1348883 lb05c11.x
39	25.4	43.8	909	48	A1568751	A1568751 th15d09.x
40	25.4	43.8	408	48	AU036448	AU036448 AU036448
41	25.4	43.8	466	51	A1717729	A1717729 UT-R-10-a
42	25.4	43.8	458	51	A1733464	A1733464 oq83h06.x
43	25.2	43.4	392	20	D31233	D31233 HML1301 Hu
44	25.2	43.4	244	20	T02622	T02622 0276C3 cbps
45	25.2	43.4	567	31	AA291389	AA291389 zt38a03.s

ALIGNMENTS

RESULT 1
AA340759
LOCUS AA340759 337 bp mRNA
DEFINITION EST146034 Fetal Kidney II Homo sapiens cDNA 5' end similar to
similar to protein phosphatase 2A, beta, 55 kDa, mRNA sequence.
ACCESSION AA340759
NTD g1992998
VERSION AA340759.1 GI:1992998

Best Local Similarity 68.5%; Pred. NO. 81;
Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

0v 5 CACACACTGGCGATGCCCATGACTCGAGTCGACATCGATTTTTTTTT 58
 ||||| | | | | | | | |
Db 257 CACACACTGGCTGTGTGAAGAACAAGCTAATTTTTTATCCTTTTTTTTTT 204

DEFINITION	AA902569	425 bp	mrna	EST	09-JUN-1998
AA902569	AA902569	425 bp	mrna	EST	09-JUN-1998
ojs9809.s1	NCI_GARP_Kid3	Homio sapiens	CDNA clone	IMAGE:1502609	3'

ACCESSION	AA902569	
NID	93037692	
VERSION	AA902569.1	GI:3037692
KEYWORDS	EST.	
SOURCE	human.	

REFERENCE	1 (bases 1 to 425)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On Jan 9, 1998 this sequence version replaced gi:936995.

Trace considered overall poor quality
Insert length: 1525 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 1

BASE COUNT	128 a.	90 c	113 g	94 t
ORIGIN				

Query Match	46.2%	Score 26.8;	DB 39;	Length 425;
Best Local Similarity	68.5%;	Pred. No. 83;		
Matches 37; Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0

Oy 5 CACACACTGCGATGCCGATGCCTGCAGTCGACATCGAATTTTTTTTTTTTTTTT 58
||||| | ||||| | | |||||
Db 92 CACACACTGGCTGGTGAGAACTCAAGGCTAATTTTTTTTATCCCTTTTTTTTTTTTTT 39

RESULT	9
LOCUS	A1026930/c
DEFINITION	A1026930 765 bp mRNA EST 27-AUG-1998
	ov82809.x1 Soares_testis_NHT Homo sapiens cDNA IMAGE:15643848

ACCESSION	AI026930	
NID	9324446	
VERSION	AI026930.1	GI:324446
KEYWORDS	EST.	
SOURCE	human.	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 765)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
On May 8, 1995 this sequence version replaced gi:1801259.

Insert length: 2811 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 466.

BASE COUNT	206 a	166 c	220 g	171 t	2 others
ORIGIN					

Query Match	46.28;	Score 26.8;	DB 41;	Length 765;
Best Local Similarity	68.58;	Pred. No. 75;		
Matches 37; Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0;

QY 5 CACACACTGGCATGCCGATGCAGTCGACATCGATTTTTTTTTTTTT 58
||||| | ||||| | | | |||||
Db 229 CACACACTGGCTGTGAGAAGCAAGGCTAAATTTTATCCTTTTTTTTTT 178

RESULT 10	LOCUS	DEFINITION
AI079505/c	AI079505 456 bp	EST 29-SEP-1998
	09955f01.x1 Soares fetal liver spleen INFES S1 Homo sapiens cDNA	
	clone IMAGE:1673553 3' similar to gb:M87503 TRANSCRIPTIONAL	

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:19:30 ; Search time 329.34 Seconds

(without alignments)
511.801 Million cell updates/sec

Title: US-09-325-095-1718.SEG

Perfect score: 1 TTTAGCTGACGATTCCTAA.....CCGACTCGACGACATCGA 53

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl3.*
10: gb_pl4.*
11: gb_pl5.*
12: gb_pl6.*
13: gb_pl7.*
14: gb_pl8.*
15: gb_pl9.*
16: gb_pl10.*
17: gb_pl11.*
18: gb_pl12.*
19: gb_pl13.*
20: gb_pl14.*
21: gb_pl15.*
22: gb_pl16.*
23: gb_pl17.*
24: gb_pl18.*
25: gb_pl19.*
26: gb_pl20.*
27: gb_pl21.*
28: gb_pl22.*
29: gb_pl23.*
30: gb_pl24.*
31: gb_pl25.*
32: gb_pl26.*
33: gb_pl27.*
34: gb_pl28.*
35: gb_pl29.*
36: gb_pl30.*
37: gb_pl31.*
38: gb_pl32.*
39: gb_pl33.*
40: gb_pl34.*
41: gb_pl35.*
42: gb_pl36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	35	66.0	35	5	A37245	A37245 Sequence 14

RESULT	1.	LOCUS	A37245	Sequence 14 from Patent WO9403609.	PAT	05-MAR-1997
DEFINITION	A37245	35 bp	DNA			
ACCESSION	A37245					
NID	g2294356					
VERSION	A37245.1	GI:2294356				
KEYWORDS						
SOURCE	unclassified					
ORGANISM	unclassified					
REFERENCE	1 (bases 1 to 35)					
AUTHORS	Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.					
TITLE	EUCARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE OR PROTEIN KINASE AND ASSAYS USING THEM					
JOURNAL	Patent: WO 9403609-A 14 17-FEB-1994.					
COMMENT	IMP CANCER RES TECH (GB)					
FEATURES	Other publication JP 8503124T 960409.					
source	1. 35					
BASE COUNT	10 a	9 c	4 g	12 t		
ORIGIN						

ALIGNMENTS

ALIGNMENTS	BOVPHOS3KN	A37245	BOVPHOS3KN
2	26	49.1	3207
3	26	49.1	3498
4	24.4	46.0	3424
5	24.4	46.0	3207
6	24.4	46.0	3207
7	24	45.3	2656
8	24	45.3	2656
9	24	45.3	75812
10	24	45.3	30469
11	23.8	43.4	15532
12	23	43.4	9829
13	22.8	43.0	3452
14	22.8	43.0	213
15	22.8	43.0	213
16	22.8	43.0	3389
17	22.8	43.0	39482
18	22.6	42.6	1702
19	22.6	42.6	15516
20	22.4	42.3	4916
21	22.2	41.9	138636
22	22.2	41.9	506
23	22	41.5	78874
24	22	41.5	152383
25	22	41.5	413
26	22	41.5	157287
27	21.8	41.1	1383
28	21.8	41.1	1418
29	21.8	41.1	687
30	21.8	41.1	590
31	21.8	41.1	27405
32	21.8	41.1	35353
33	21.8	41.1	612
34	21.8	41.1	612
35	21.6	40.8	38078
36	21.6	40.8	231305
37	21.6	40.8	42734
38	21.6	40.8	32911
39	21.4	40.4	99384
40	21.4	40.4	5417
41	21.4	40.4	78651
42	21.4	40.4	3213
43	21.4	40.4	110000
44	21.4	40.4	60649
45	21.4	40.4	86884

Query Match 66.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTAAGCTTAAGCTTAAGCTATCATCC 35
|||||
DB 1 TTTAAGCTTAAGCTTAAGCTATCATCC 35

RESULT 2
BOVPHOS3K/c

LOCUS BOVPHOS3K 3207 bp mRNA MAM 19-AUG-1992
DEFINITION Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete cds.

ACCESSION M93252
NID G163519
VERSION M93252.1 GI:163519
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE Bos taurus cDNA to mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Bovidae; Bovinae; Bos

REFERENCE 1 (bases 1 to 3207)
Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R.,
Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F.,
Hsuan, J.J., Courtneidge, S.A., Parker, P.J. and Waterfield, M.D.

TITLE Phosphatidylinositol 3-kinase: Structure and expression of the 110
kd catalytic subunit
JOURNAL Cell 70, 419-429 (1992)
MEDLINE 92354059
FEATURES

source location/Qualifiers
1..3207
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBR-1"
1..3207
/EC_number="2.7.1.67"
/product="phosphatidylinositol 3-kinase"
/protein_id="AAA30698.1"
/db_xref="PID:g163520"
/db_xref="GI:163520"
1..3207
/translation="MPPRSSGELMGHLMPPRIIVECLIPNGMIVTLECLREATLIT
IKHELFKARKYPLHQLODESSYIFSVTOEAREEFDETRRLCDLRFQPLKVI
EPVGRREKILNREIGFAIGMPCEDMDKDEVDPRNINLVCKEAVDLRLDSPH
SRAMVYVPNVSSPELPKHYNKLDKGLIYIVIVISPNNDKOKYTLKINHCVPE
OYIAEAIKRTKRSMLSSBOLKCYLEVOGRTIILKVCCEDEFLKRYPLSQYKTRSC
IMLGMPNLMIMAKESLSQLPMDCTMPSTSRISTATPYMNETSTKSLMVISAL
RIKICATYVNVNINIDIKIYRTGIIYHGSEPLCDNVATQVPCSNPKNMELNVDIY
IPDLPRARLCLISICSVGRKAEKHEOPLANGNINLFEDYDTIVSGKMLNLPVPH
GLEDLNPIGVTSNPKETPCLEPEPFMSVYKPEPMYSIEEHANVSREGEFSY
SHAGISNLRADNENRENDKQLRAICRDRPLSETTDEKPLMSHRHVCYTIPEILP
KULSVKNSNDEVAQMTCLYKMPPIPEPDAEMELDQNYDPNVRGFAVCELYTL
DDKLSQYLIQLVQVLYEQYLDNLVRLKALNQRIGHEFFHNLSEHNKTVSQ
RFGILLESYCRACGMVTLKHLNROYEAMKILNLDILKOEKDETOKVMKFLVQNR
RPDMFALOGFSLPNPAHQGNLRLKECRIMSAKRLMNMENPDIMSELHQNNE
IIFKNGDRLDMGLTIQIRIMENIMONOGDLNMLPYGCLISIGCVGLIEVRSHT
IMOIQCGKGLGALQPNSTHLOMKDKNGEYDAADITPSCAGCVATFTLLIG
DRHNSNIVKRDGQLFHTDFGHFLDHKKKRYKREVPFVLTODFLIVISGAOECT
KTRFEREOMCYKAYLAIRQHANLFINLFESMLGSGMPELQSPFDIAYIRKTLALDK
TEQEALEFEMOMNDAHHGWTTKMDWTFHTIKOHALN"

BASE COUNT 1028 a 581 c 680 g 918 t
ORIGIN

Query Match 49.1%; Score 26; DB 3; Length 3207;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AGCATTCCTAAGCTCATCATCC 35

DB 110 AGCATTCCTAAGCTCATCATCC 85
|||||

RESULT 3

LOCUS A37232 3498 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9403609.
ACCESSION A37232
NID G2294345
VERSION A37232.1 GI:2294345

KEYWORDS unidentified.
SOURCE unidentified.

REFERENCE 1 (bases 1 to 3498)
Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
EUCARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL IMP CANCER RES TECH (GB)
Patent: WO 9403609-A 1 17-FEB-1994;
Other publication JP 8503124T 960409.

COMMENT location/Qualifiers
source

CDS

1..3498
/organism="unidentified"
/db_xref="taxon:32644"
1..3207
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PID:e304808"
/db_xref="PID:g2294346"
/db_xref="GI:2294346"

1..3207
/translation="MPPRSSGELMGHLMPPRIIVECLIPNGMIVTLECLREATLIT
IKHELFKARKYPLHQLODESSYIFSVTOEAREEFDETRRLCDLRFQPLKVI
EPVGRREKILNREIGFAIGMPCEDMDKDEVDPRNINLVCKEAVDLRLDSPH
SRAMVYVPNVSSPELPKHYNKLDKGLIYIVIVISPNNDKOKYTLKINHCVPE
OYIAEAIKRTKRSMLSSBOLKCYLEVOGRTIILKVCCEDEFLKRYPLSQYKTRSC
IMLGMPNLMIMAKESLSQLPMDCTMPSTSRISTATPYMNETSTKSLMVISAL
RIKICATYVNVNINIDIKIYRTGIIYHGSEPLCDNVATQVPCSNPKNMELNVDIY
IPDLPRARLCLISICSVGRKAEKHEOPLANGNINLFEDYDTIVSGKMLNLPVPH
GLEDLNPIGVTSNPKETPCLEPEPFMSVYKPEPMYSIEEHANVSREGEFSY
SHAGISNLRADNENRENDKQLRAICRDRPLSETTDEKPLMSHRHVCYTIPEILP
KULSVKNSNDEVAQMTCLYKMPPIPEPDAEMELDQNYDPNVRGFAVCELYTL
DDKLSQYLIQLVQVLYEQYLDNLVRLKALNQRIGHEFFHNLSEHNKTVSQ
RFGILLESYCRACGMVTLKHLNROYEAMKILNLDILKOEKDETOKVMKFLVQNR
RPDMFALOGFSLPNPAHQGNLRLKECRIMSAKRLMNMENPDIMSELHQNNE
IIFKNGDRLDMGLTIQIRIMENIMONOGDLNMLPYGCLISIGCVGLIEVRSHT
IMOIQCGKGLGALQPNSTHLOMKDKNGEYDAADITPSCAGCVATFTLLIG
DRHNSNIVKRDGQLFHTDFGHFLDHKKKRYKREVPFVLTODFLIVISGAOECT
KTRFEREOMCYKAYLAIRQHANLFINLFESMLGSGMPELQSPFDIAYIRKTLALDK
TEQEALEFEMOMNDAHHGWTTKMDWTFHTIKOHALN"

BASE COUNT 1147 a 626 c 722 g 1003 t
ORIGIN

Query Match 49.1%; Score 26; DB 5; Length 3498;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AGCATTCCTAAGCTCATCATCC 85

RESULT 4
LOCUS HSPH13K 3424 bp mRNA PRI 24-AUG-1995
DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
ACCESSION Z29090
NID 9472990
VERSION Z29090.1 GI:472990
KEYWORDS phosphatidylinositol 3-kinase.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 1068)
TITLE Volinia, S., Hiles, I., Ormondroyd, E., Mizetich, D., Antonacci, R.,
Rocch, M. and Waterfield, M.D.
JOURNAL Molecular cloning, cDNA sequence, and chromosomal localization of
MEDLINE the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene
REFERENCE 95229146
AUTHORS 2 (bases 1 to 3424)
TITLE Volinia, S.
JOURNAL Direct Submission
TITLE Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig
JOURNAL Institute for Cancer, Research, 91 Riding House Street, London, W1P
8BT, UK
FEATURES
Source Location/Qualifiers
1..3424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="K562"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CAA82333.1"
/db_xref="PIR:G472991"
/db_xref="GI:472991"
/db_xref="SWISS-PROT:P42336"
translation="MPPRSSGELMGHIMPPRLIVCELPNGMIVTECLREAVLT
IKHLEKARKYPLHQLQDSSSTIFVSVQEARREFFEDTRLCDFRLQPLKVI
EPVGNREKILNREIGALIGPVCPEPMVNDPEVQDRNRNLNVCRAVDLRDLSH
SRNAVYPPNYESPELPKHYNKLDGQIIVIVIVSPNNDKQYTLKINHCVSE
QVIAEAIKRTKRLSLSEDLKLCVLEOGYILKVCCEDEYFLEKYPILSOYKIRSC
IMGRMNLIMAKESLYSOLPMDCFMPSYSRRISTATPMNGESTKSLIMVNL
RIKILCATYVNLSDIKTYVRGIYHGGEPICLDNNTQVQVPSNRNMLNWDIT
IPDIPRARCLSLSCYGRKGAKEHCPLAMGNINLDTYDITLVSGKMLNMPVH
GLEDLNPICVTSNPKPEPCLLEEDPSSVYKPDMSYIEHANSVREGFSY
SHAGLSNRLANDENLRNDKQALISTROPLESTEQEDFLSHRYCYTILEILP
KLLSVKNSRDEVAQMYCLVMDPPIKPEQAMLLDQNPDPVAVGAVCLERKYL
DKLSOYLIOVYLYEYLDNLVFLKALTNORIGHFFFMHLSKEMHNTVSO
RPGILLESYRACGMYLKHLNRYEAMEKILNLDIKOEKDETQYOMKYLEONR
RPDMDLQGLSLNPAHOLGNLKECRIMSSAKRPLMNMENPDIMSELFOHNE
IIFKNGDGLQDMITLQIRIMENIMQNGIDMLPYGCLSIDCVGLIEVYNSHT
IMQIOCGGLKGLQFNSHTLHQLKDKNKEIYDAIDLETSCAGCAVATFLIGT
DRHNSINIMKDDGLFHDIFGHFLDKKKKRGYRERPEVLTQDFLIVISKGOECT
KTRFEREEMCYRAYLAIRHNLFINLSMMSGMPELQSFDDIAYIRKTLALDK
TEQALFYFMKQMDAHGGWTKMDWIFHTIKQHALN"

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 3207)
TITLE Stridivant, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledder, L.M.,
O'Leary, A. and Heimbrock, D.C.
JOURNAL Catalytic Activity of the p110-alpha Subunit of Human
MEDLINE Phosphoinositide 3'-Hydroxylase is Required for Signal
REFERENCE 9414994
AUTHORS 2 (bases 1 to 3207)
TITLE Stridivant, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledder, L.M.,
O'Leary, A. and Heimbrock, D.C.
JOURNAL Direct Submission
TITLE Submitted (21-NOV-1996) Cancer Research, Merck Research Labs,
JOURNAL Summeytown Pike, West Point, PA 19486, USA
FEATURES
Source Location/Qualifiers
1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3207
/function="phospholipid kinase"
/codon_start=1
/product="phosphoinositide 3'-hydroxylase p110-alpha
subunit"
/protein_id="AAB39753.1"
/db_xref="PIR:G1763626"
/db_xref="GI:1763626"
translation="MPPRSSGELMGHIMPPRLIVCELPNGMIVTECLREAVLT
IKHLEKARKYPLHQLQDSSSTIFVSVQEARREFFEDTRLCDFRLQPLKVI
EPVGNREKILNREIGALIGPVCPEPMVNDPEVQDRNRNLNVCRAVDLRDLSH
SRNAVYPPNYESPELPKHYNKLDGQIIVIVIVSPNNDKQYTLKINHCVSE
QVIAEAIKRTKRLSLSEDLKLCVLEOGYILKVCCEDEYFLEKYPILSOYKIRSC
IMGRMNLIMAKESLYSOLPMDCFMPSYSRRISTATPMNGESTKSLIMVNL
RIKILCATYVNLSDIKTYVRGIYHGGEPICLDNNTQVQVPSNRNMLNWDIT
IPDIPRARCLSLSCYGRKGAKEHCPLAMGNINLDTYDITLVSGKMLNMPVH
GLEDLNPICVTSNPKPEPCLLEEDPSSVYKPDMSYIEHANSVREGFSY
SHAGLSNRLANDENLRNDKQALISTROPLESTEQEDFLSHRYCYTILEILP
KLLSVKNSRDEVAQMYCLVMDPPIKPEQAMLLDQNPDPVAVGAVCLERKYL
DKLSOYLIOVYLYEYLDNLVFLKALTNORIGHFFFMHLSKEMHNTVSO
RPGILLESYRACGMYLKHLNRYEAMEKILNLDIKOEKDETQYOMKYLEONR
RPDMDLQGLSLNPAHOLGNLKECRIMSSAKRPLMNMENPDIMSELFOHNE
IIFKNGDGLQDMITLQIRIMENIMQNGIDMLPYGCLSIDCVGLIEVYNSHT
IMQIOCGGLKGLQFNSHTLHQLKDKNKEIYDAIDLETSCAGCAVATFLIGT
DRHNSINIMKDDGLFHDIFGHFLDKKKKRGYRERPEVLTQDFLIVISKGOECT
KTRFEREEMCYRAYLAIRHNLFINLSMMSGMPELQSFDDIAYIRKTLALDK
TEQALFYFMKQMDAHGGWTKMDWIFHTIKQHALN"

AUTHORS Klippel, A., Escobedo, J.A., Hirano, M., and Williams, L.T.
TITLE The interaction of small domains between the subunits of phosphatidylinositol 3-kinase determines enzyme activity
JOURNAL Mol. Cell. Biol. 14, 2675-2685 (1994)
REFERENCE 94187738
AUTHORS 2 (bases 1 to 3207)
TITLE Klippel-Giese, A.
JOURNAL Direct Submission
 Submitted (05-NOV-1993) Anke Klippel-Giese, CVRI and Dalilhi Research Center, University of California San Francisco, 505 Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
FEATURES Location/Qualifiers
 1..3207
 /organism="Mus musculus"
 /strain="Balb/c"
 /db_xref="taxon:10090"
 1..3207
 /standard_name="p110 protein"
 /codon_start=1
 /product="phosphatidylinositol 3-kinase 110 kDa subunit"
 /protein_id="AA18334.1"
 /db_xref="PID:9414995"
 /db_xref="GI:414995"
 /translation="MPPPSGSELMGILHMPRIIVCELLPNGMYTLECLREATLV
 IKHELFEARKYPLHQLQDETSYIFSVTORAREPERDETRRLCDLRLPQPLKVI
 EPGNREKILNREIGFVGMPCEDPMWKDEBQDFRNILNCKEANDIDNSPH
 SPAAVYPPNVSPELPKHTINKDKQIIVIVIVTSPNDKQITIKINDCVPE
 QYIAEIRKTRSMILSSQLKCYLEYQKIVLVKCGDEYFLEKYSQYKIRSC
 IMIGRPMIMLAKESLYSQLIDSETPMSRSRISTATPAPCSNPNEMLVNSAL
 RKIKICATVYNNIRIDIKIVYRGCIHGSGELCDVNPAPCSNPNEMLVNSAL
 IPDLPRILALCISGVKGRKAKHEHCLANGNINLPYTLTYISGKALMLPVPY
 GLEDLNPICVTSNPKETPELETFPMSVYVPEKMSVIEHANSVSEAGPSY
 SITGSNRLARQNELENDKEDRLACTDPSEITBOKRDLNRRHVCVITPELP
 KILLSKMSRDEVAQMTCLVDWPPKPEQAMELDCVTPPPMRSKSEVRLLEYLT
 DKLSQYLLIOLVQVLYKEQYLDNLVRLFKALNQRIGHFFFMHAKEMNKVYQ
 RGLLSEYCGACGMLKLNROVEAMEKLTDLKOEKDEQKVMKELVQMR
 QDPMDALQGFSLPMHQLGNLRLEGRINSKARPEMLNEMNDIMSELFPQMR
 IIFKNGDRLQDMITLQIRITIMINMONGDLPMIPGCSLTGCGVGLTVRSHT
 IMQIOCKGLKALQENSHLHOMLKDKRKGEITYDAIDLFRSCAGCVANFTIIG
 DHNSNINWKDGLPHIDEGHFLDKRKRFYKRRRVFVLTDQFLIVISGAEYGT
 KTRFERQEMCYKAYLAIRQANLEFINLSMGLSGMPELQSPDIAYIRRTLADK
 TEOEALEYETKQMDAHGCGTMMKMWIHTIKQALN"

BASE COUNT 958 a 693 c 734 g 822 t
UNIGIN

Query Match 46.0%; Score 24.4; DB 12; Length 3207;
Best Local Similarity 96.2%; Pred. No. 5.5;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 AGGCATCTAAAGTCACTATCATCCC 35
 |||||||||||||||||||||
Db 110 AGGCATCTAAAGTCACTATCATCCC 85

RESULT 7
MATGENB/c MATGENB 166924 bp DNA circular PLN 13-NOV-1998
LOCUS A.thaliana mitochondrial genome, part B.
DEFINITION Y08502
ACCESSION 91785729
NID Y08502.1 GI:1785729
VERSION 18S ribosomal RNA; 26S ribosomal RNA; 5S ribosomal RNA; apocytochrome b; atp1 gene; atp6 gene; atp9 gene; ATPase subunit 1; ATPase subunit 6; ATPase subunit 9; cb203 gene; cb206 gene; cb256 gene; cb382 gene; cb452 gene; cob gene; cox1 gene; cox2 gene; cox3 gene; cytochrome c biogenesis orf203; cytochrome c biogenesis orf206; cytochrome c biogenesis orf256; cytochrome c biogenesis orf382; cytochrome c biogenesis orf452; cytochrome c oxidase subunit 1; cytochrome c oxidase subunit 2; cytochrome c oxidase subunit 3; mat1 gene; maturase; nad1 gene; nad2 gene; nad3 gene; nad4 gene; nad4l gene; nad5 gene; nad6 gene; nad7 gene; nad9 gene; NADH dehydrogenase subunit 1; NADH dehydrogenase subunit 2;

SOURCE ORGANISM
 Mitochondrion Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; rosidae; Caprales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 166924)
 Unsel, M., Marlenfeld, J.R., Brandt, P. and Brennicke, A.
 The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366924 nucleotides
 Nat. Genet. in press
AUTHORS 2 (bases 1 to 166924)
 Marlenfeld, J.R.
TITLE Direct Submission
 Submitted (30-SEP-1996) J.R. Marlenfeld, Institut fuer Genbiologische Forschung GmbH, Innestrasse 63, 14195 Berlin, FRG
JOURNAL 3 (bases 1 to 166924)
 Marlenfeld, J., Unsel, M., Brandt, P. and Brennicke, A.
AUTHORS Genomic recombination of the mitochondrial atp6 gene in Arabidopsis thaliana at the protein processing site creates two different presequences
REFERENCE 97191539
JOURNAL DNA Res. 3 (5), 287-290 (1996)
COMMENT Related sequences:
 X96535, X96536, X91201-X91210, X72616, X67736, X65123, X67105, X60045, X60047, X60048, M58320, X07376, X94583, X82565, U49402 & U49403.
FEATURES SOURCE
 1..166924
 /organism="Arabidopsis thaliana"
 /mitochondrion
 /variety="Columbia"
 /db_xref="taxon:3702"
 /clone.lib="lorist x"
 /sub_clone="puc19"
 /sub_clone="puc18"
 /sub_clone="pbluescript SK"
 /sub_clone="pbluescript KS"
 1729..2097
 /gene="nad1"
 /note="orf122c"
 /codon_start=1
 /protein_id="CA65844.1"
 /db_xref="PID:e283572"
 /db_xref="PID:91785730"
 /db_xref="GI:1785730"
 /db_xref="SPRMBL:P92510"
 /translation="MKHASFCLSRILLAPCRILGTLILLPFGCTLRQFLIARS
 LFIRHVEWIPGSLHSHHPFPLASPPLVETLLIARLLSLIKQLSLPAPKASIASL
 DAKTRGSLSSCSYCYPT"
 complement(3588..4310)
 /gene="nad1"
 /note="orf140a"
 /codon_start=1
 /protein_id="CA65844.1"
 /db_xref="PID:e283453"
 /db_xref="PID:91785731"
 /db_xref="GI:1785731"

```

/db_xref="SPTREMBL:P92511"
/translation="MRSYVLRSLRGRLVIVLESTRKRLRSTNIVPRKKGOKSKK
NMARKGAPILVLRGKNSSDSRSAELJCLTFEYIVAPTLGVLFLRLHFWM
GLRLGDNENHFVGVDGATGLDLNPQEQPTIGVNRALDLNELPLVHLVE
VEPOSTKQANDWLAHLNQVQLTRDLQTEPNIMRQALIDIDMEVRSIQRRFIE
RORDRLREVORSWLREQLNRYR"
complement(4340..4412)
/gene="trna-lys"
/product="transfer RNA-Lys"
complement(4340..4412)
/gene="trna-lys"
complement(7553..7915)
/gene="nad1"
/note="orf120"
/codon_start=1
/protein_id="CA69816.1"
/db_xref="PID:e283453"
/db_xref="PID:g1785732"
/db_xref="GI:1785733"
/db_xref="SPTREMBL:P92512"
/translation="MNRITIEKVSMLCEGCLPTEPRADAANTAVHIINKYSTAINF
HVDENVFQGVPTVYLRRGCVAYIHCDEGKIKLPRAKGSEKSYLINRIVSLTYI
GIGKTSPPRAKSHLGIRG"
9499..9822
/gene="nad1"
/note="orf107d"
/codon_start=1
/protein_id="CA69817.1"
/db_xref="PID:e283573"
/db_xref="PID:g1785733"
/db_xref="GI:1785733"
/db_xref="SPTREMBL:P92513"
/translation="MEDEGFTALRSVYIVVERISSSLSYNTLSSATRLDMWIKT
PRHRAAPRALGIONQICTACPHLSALFGLSVASCPSKNDGLHTRIFYDFEFP
DSAVF"
17631..17639
/gene="cox3"
/evidence="not_experimental"
17631..19077
/gene="cox3"
18280..19077
/gene="cox3"
/codon_start=1
/evidence="not_experimental"
/product="cytochrome c oxidase subunit 3"
/protein_id="CA69818.1"
/db_xref="PID:e283454"
/db_xref="PID:g1785734"
/db_xref="GI:1785734"
/db_xref="SWISS-PROT:P92514"
/translation="MISQSRSHYLVDPSPWPISGJGALATTVGVMYMHFPGGAR
LTSIGLILFITYMEVWRDYLRETLLEGHHTKVVQLPRGSLIFIVSEVFFPAFW
ASSHSLAPAVEIGIWPKGIENVLDPMEIPFLNTPILPSGAIVTAHHHILGKXK
RAYVALAVIALVLTGFGMEYQAPFIIISDIYSTFLATGFGFHHYIIGTLLE
ITGIRYLIHLTLEHNVFEAANMYHFPDVVWLFVSLIYMGGI"
complement(20471..20773)
/gene="nad1"
/note="orf100a"
/codon_start=1
/protein_id="CA69819.1"
/db_xref="PID:e283574"
/db_xref="PID:g1785735"
/db_xref="GI:1785735"
/db_xref="SPTREMBL:P92515"
/translation="MLHISOEFLPISMEHRMLITSOVIKKIGIPKVIHFARONN
RGILFQIFHHIDLEFGALIKYVVLDKSRQVINDSLINSLVLYXNQKLEK"
20830..21189
/gene="nad1"
/note="orf119"
/codon_start=1
/protein_id="CA69820.1"
/db_xref="PID:e283455"
/db_xref="PID:g1785736"

```

```

/db_xref="GI:1785736"
/db_xref="SPTREMBL:P92516"
/translation="MHRKSEGHSLPLCMHRSMODISCHLQMPREVLQAGFYPT
TFIDAHFVSCDACORKNFTIRNEPQHFILVEVDFWGIYFNKRTLFQWKPIDH
NGRGLCKMGSSCEPH"
complement(21700..22029)
/gene="nad1"
/note="orf109b"
/codon_start=1
/protein_id="CA69821.1"
/db_xref="PID:e283456"
/db_xref="PID:g1785737"
/db_xref="GI:1785737"
/db_xref="SPTREMBL:P92517"
/translation="MLRSKLELEVVADKDEEQVYVRKAKQIVNPELMLSSARIG
LRWPFYKCSPLTSOSTAPYGTSPMRGKHLDLNFGNNLSLEIRSSYTSFTIN
SLGRFL"
complement(22687..22989)
/gene="nad1"
/note="orf100b"
/codon_start=1
/protein_id="CA69822.1"
/db_xref="PID:e283575"
/db_xref="PID:g1785738"
/db_xref="GI:1785738"
/db_xref="SPTREMBL:P92518"
/translation="MGLSTCOLFESKRPNNTNHSVPMAGSLKSYPPSNINITSCV
SIXINDRNLYLDEGRVRLSLINIGERCDEVPELVSVELFPLQSSGYEKLEHSEN"
27087..27619
/gene="nad1"
/note="repeat 11"
27109..27196
/gene="trna-ser"
/product="transfer RNA-Ser"
27109..96784
/gene="trna-ser"
27502..27577
/gene="trna-tyr"

Query Match 45.38; Score 24; DB 7; Length 166924;
Best Local Similarity 75.08; Pred. No. 7.2;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10 AGGCATTCTAAGTCATCATCCCGACTCGAGTCGACA 49
DB 58270 AGGTATCTCTAAATCAAGATATCCCGACGTGAGAGCTCA 58231

RESULT 8
SCYOR007C 2656 bp DNA PLN 07-AUG-1997
LOCUS S.cerevisiae chromosome XV reading frame ORF YOR007C.
DEFINITION 274915 y13140
ACCESSION 274915 y13140
VERSION 274915.1 GI:1420099
KEYWORDS
SOURCE
ORGANISM baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 2656)
AUTHORS Pettersson,B., Sterky,F. and Uhlen,M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2656)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-1996) Data collected by MIPS on behalf of the
European yeast chromosome XV sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
FEATURES
location/Qualifiers
1..2656
/organism="Saccharomyces cerevisiae"

```

/db_xref="taxon:4932"
 /chromosome="XV"
 complement(317..1357)
 /note="ORF YOR007c"
 /protein_id="CA69195.1"
 /db_xref="PID:e251943"
 /db_xref="PID:g1420100"
 /db_xref="GI:1420100"
 /db_xref="SPTREMBL:Q12118"
 /translation="MSASKEEIALIVNFSSIVEKEISEGADSLVAVNDICSEAF
 EGERAVSCILKSEFGQHLADILNSASRVPSNKKDAENVEIPIPDADYTKA
 EDLKGKGNKAMANKDELAINTYTAIRYLPNAYARAASHSIEKRYDAVDAAE
 VASISDPSYFRKYSRIGFAKYAGKREALEAKRYLDIEGNATAMKRDYSAKK
 VQSLNLEKTVPEQSDADVDAQSAGSLPDVGLGGLGGLGGLMNPOLMQAOKM
 MSNPGAMQIKQKMDPSIRQMAEGFASGGCTPNISDLMMNPALRMAGNLFQAGAQ
 STDETPDNENK"
 complement(1678..1750)
 /gene="tr(ACG)OR - systematic name"
 /note="anticodon gene: ACG; tRNA-Arg2 - common name"
 complement(1678..1750)
 /gene="tr(ACG)OR - systematic name"
 complement(1949..2273)
 /note="delta"
 BASE COUNT 819 a 491 c 426 g 920 t
 ORIGIN

Query Match

Best Local Similarity 45.3%; Score 24; DB 7; Length 2656;
 Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 4 AACCTAGCATTCCTAAAGTCAATCATCCGACCTCGAGTCGACATC 51
 DB 602 AAGATCTGCATATCCACCCGACCTAGCAGCTTGCATCGACATC 649

RESULT 9
 LOCUS ATAC007143 75812 bp DNA PLN 05-JUN-1999
 DEFINITION Arabidopsis thaliana chromosome II BAC T17H1 genomic sequence,
 complete sequence.
 ACCESSION AC007143
 NID 95001456
 VERSION AC007143.3 GI:5001456
 KEYWORDS HTG.
 SOURCE Thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsids.
 REFERENCE 1 (bases 1 to 75812)
 AUTHORS Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
 Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M., Benito,M.,
 Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,
 Fraser,C.M. and Venter,J.C.
 Arabidopsis thaliana chromosome II BAC T17H1 genomic sequence
 Unpublished
 2 (bases 1 to 75812)
 Lin,X. and Kaul,S.
 Direct Submission
 Submitted (23-MAR-1999) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
 3 (bases 1 to 75812)
 Lin,X.
 Direct Submission
 Submitted (05-JUN-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Jun 5, 1999 this sequence version replaced g1:4512639.
 Address all correspondence to:
 Xiaoying Lin
 The Institute for Genomic Research

9712 Medical Center Dr.
 Rockville, MD 20850, USA
 e-mail: xlin@tigr.org
 BAC clone T17H1 is from Arabidopsis chromosome II and is near the
 molecular marker m1421.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.

Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL (available by anonymous ftp
 from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
 Washington), GenScan (Chris Burge,
 http://genome.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
 (http://www.cds.dtu.dk/netgene/cbsnetgene.html), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (http://www.tigr.org/tdb/at.html).

Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown proteins'.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
 http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
 identified by repeatmasker (Arian Smit,
 http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
 genomic sequence that are not annotated as genes but have predicted
 exons by GRAIL are annotated as misc features.

The entire sequence of T17H1 is nearly identical to the published
 mitochondrial genome. Since we cannot be certain that the genes
 found in this BAC are expressed from the nucleus, we have chosen
 not to annotate this BAC.

FEATURES

source

Location/Qualifiers
 1..75812

/organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="II"
 /map="m1421"
 /clone="T17H1"

misc_feature

complement(1..5719)
 /note="Overlap with BAC clone T5M2 (1..5718)."
 complement(1..51679)
 /note="Overlap with BAC clone T5M2 (1..51675)."
 complement(1..56697)

misc_feature

/note="matching Arabidopsis ecotype C24 mitochondrial
 genome, part B (GB:Y08502:17688..74459)."
 complement(58488..66262)

misc_feature

/note="matching Arabidopsis ecotype C24 mitochondrial
 genome, part B (GB:Y08502:9924..17696)."
 complement(66259..75812)

misc_feature

/note="matching Arabidopsis ecotype C24 mitochondrial
 genome, part B (GB:Y08502:134052..143607)."
 complement(134052..143607)

BASE COUNT

20824 a 17017 c 16566 g 21405 t

ORIGIN

Query Match 45.3%; Score 24; DB 8; Length 75812;
 Best Local Similarity 75.0%; Pred. No. 7.3;
 Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10 AAGCATTCCTAAAGTCAATCATCCGACCTCGAGTCGACA 49
 DB 16203 AAGTATTCCTAAAGTCAATCATCCGACCTCGAGTCGACA 16242

RESULT 10
 LOCUS SCU43491 30469 bp DNA PLN 25-OCT-1996
 DEFINITION Saccharomyces cerevisiae cosmid clone pEOA156 from chromosome XV.
 ACCESSION U43491 X95109
 NID 91150992

VERSION 043491.1 GI:1150992
KEYWORDS Baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 30469)
AUTHORS Sterky,F., Holmberg,A., Pettersson,B. and Uhlen,M.
TITLE The sequence of a 30 kb fragment on the left arm of chromosome XV
from Saccharomyces cerevisiae reveals 15 open reading frames, five
of which correspond to previously identified genes
JOURNAL Yeast 12 (108 suppl), 1091-1095 (1996)
MEDLINE 97051599
REFERENCE 2 (bases 1 to 30469)
AUTHORS Sterky,F. and Uhlen,M.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1995) Fredrik Sterky, Department of Biochemistry
and Biotechnology, Royal Institute of Technology, Teknikringen 34,
Stockholm, 100 44, Sweden
COMMENT On Apr 10, 1996 this sequence version replaced gi:1237089.
FEATURES
source
Location/Qualifiers
1..30469
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="XV"
/clone="cosmid clone pEOA156"
/gene="UNB378"
complement(127..1263)
/gene="UNB378"
complement(127..1263)
/gene="UNB378"
note="ORF UNB378, unidentified function"
/codon_start=1
/product="hypothetical protein UNB378"
/protein_id="AAC49477.1"
/db_xref="PID:g1150993"
/db_xref="GI:1150993"
/translation="MPYKLRPMGIAIPFLISFTIGYGAHYFLSNFLSPKQITE
FCLSMITSLYLAICTNGRPLPYKPPDWMNFCKCOSYKPERSHCHTCQCVL
MMDHCHWIMNCVGFANYPHFLRFLEIIVTTSVLCIQARIYFIMQQRILPGYFK
KSELIPLISSLFVLLITLILFLCLFNOILNGSISMDRLSLESLFNGRLT
OKLIDNTWRIYPERSRFONKKDAEHLTKRPREDELVNPEDYDELTYNALLYGPILH
LMILMPYGVPTGDGNNPFRKNGISKEANSISLEDHLISLPMPDGGKTNTVFNHGSSTLE
MRNDSGQILRTPLPONGRHASREKWNINDGESLDDGVVDVME"
complement(1680..2663)
/gene="UNB327"
complement(1680..2663)
/gene="UNB327"
note="ORF UNB327, unidentified function"
/codon_start=1
/product="hypothetical protein UNB327"
/protein_id="AAC49478.1"
/db_xref="PID:g1150994"
/db_xref="GI:1150994"
/translation="MLHAGESFOMSTLLERTKSVOLKRRAAKTSANPAEYAKAK
VLRILYSMDPEPWRNDPLIHGYETSEFIFTSPLTINNESVNTISHLPLALG
FTVLLDKSTIKVFAITWLDHNVIDLFTSGAFACILSSSFQLSHSLRATLGN
KILYLGLICILIVSMVAILLYGYEPEKSLFALILVSFISACISVLSKDKFRKRM
RYPAGLVEFGGLSIIPIFGILCYFSEFWIQLQFVLLVGLVILGVALYGMF
PERICPKFDMGSHQLEHFLVIVIALCHLRLGLNSYELVHMGNGIYS"
3518..4339
/gene="PHO80"
/db_xref="SGD:S0005361"
3518..4339
/gene="PHO80"
/note="Pho80"
/codon_start=1
/db_xref="SGD:S0005361"
/product="phosphate system cyclin"
/protein_id="AAC49479.1"
/db_xref="PID:g1150995"
/db_xref="GI:1150995"
/translation="MESTSGRSENHEDGIPKVIILPADFNKSKRTDIIVLISRMVY
SLAINENSATKKSDDQITILRIYHASKIPNISINIVYIRLTKKSLCHCYLMTSLYII

DILQTVPDFTLNLSLTHRELLTATTATAGLDSSESTNAHYAVGVGRHELTLEN
DLKRVNRIILPRDNNITLCSIEQKKKPVLDKNAISLSDISTSYNRRPSGYNVD
KYRRIIVQLVGSFNASPDKSRKYDVLPPIIDIVSEGSQTOLKSSSPNHSQKR
YSEAKDAHIYKRSKPD"
5101..7302
/gene="UNC733"
5101..7302
/gene="UNC733"
note="ORF UNC733, unidentified function"
/codon_start=1
/product="hypothetical protein UNC733"
/protein_id="AAC49480.1"
/db_xref="PID:g1150996"
/db_xref="GI:1150996"
/translation="MTSENPVILSRVINYVRAASSIASODVDFYKMLDRGFSKDLK
KADRLAMANEIISIDENHSEFLKEDSIDLWNNGNIDNLEMSDLSLDMKA
INSGRSDLOVGEFSKNSPKRREKOLKRSPIIDNSSEPFILKREPNALK
PLSLSLVLDDENNPESHYPIYETIEDHDSPEILQIREISKSDSVPLVDT
STLESLELDKNTKEIYAVDLEHDSYGVGLMDISTREYLDVTLKRENLAI
LNEVTPNSIVKVFHGAFMIDIMQRLDGLVGLFDTYHASKAIGLRSLALLEN
FANEKTSKRYQLADMRIPLSKPMTAYARADTHEILMIYQLRMLIESNKLAVLE
SRNVAKRREYSKRYPLTSSSEYSPLEKSPKILMYOINIPEREVLVRELQMD
LIARDESPREVPNOLLAIVAYPTDVIYGVSLNGVTEHYRONAKILIANIRRA
LRNKTNNNEATPSPSEETKADGILLETISVPOIRDMERSVYCNISKSRKPT
NSLTLKILPREHDTAISDGLPNKVKTEDITRAONKRSALANLEDILFELEKPL
VVPVKEIKITVDPASAPNHSPEIDNDDLVILKKNIQKQPKKEGVLEKDAVDS
KIPNIIISNKGQNNRQKKRRRFPSSSDSNCPRAKKRRRAKKNLSFR"
7686..9320
/gene="UNA544"
/db_xref="SGD:S0005528"
7686..9320
/gene="UNA544"
note="ORF UNA544, unidentified function"
/codon_start=1
/db_xref="SGD:S0005528"
/product="hypothetical protein UNA544"
/protein_id="AAC49481.1"
/db_xref="PID:g1150997"
/db_xref="GI:1150997"
/translation="MAIGKRLVLRKPAESFVAPMYDPLYPFRVGNOMLPEYIIFV
CAVLRCTIGLIGPSGSPKIDGPEFAOHMMETIHOILPSKKWYMDIOLWGDVYP
LTFHSLILGLIGSFENPSPFALKESGSPDNGLKTVKRSYIISDILPPAVIV
FTWLGIRYNQSPITGOSIASAILIQSMLIDBHGVOYNSVMGLTAAVAINLNDL
YANAACFVLISICKOMALYAPAFEFYALISLFEKFNARLITAFATLAFALII
FAPLYFGLGLKNIHOCILHRIFFPARGILPDKVNFQCVNVEYKEREPTLOLQAI
SLIATVIGFLPAMIMTLHPKHLIPVLVLAGSMSFELSFQVEYKILLPLDITLL
YSSIDMNVLSLWINNVALLFTLMLPKKGLHGOVAVSLLSMLGNSFETPREL
PKSLTPSPSISINSIDRRSLDPIYNNVKSFIITGYIAGFTHELDQFAPPSKTPD
LWVILNCAVGFICFSIFLWMSYKIFITGSKSMKDL"
9724..11160
/codon_start=1
/product="subtilisin-like protease III precursor"
/protein_id="AAC49482.1"
/db_xref="PID:g1150998"
/db_xref="GI:1150998"
/translation="MKSFTLIPILMANCCIMTIPDFGIVRFIENIDGTRSVRAGG
LGQDHPGNFTEHGVNAHKEFLPYRYVIFENEDISQIQSHQVVOXHSFVGL
TENDAFRVISSVSYSKSGQIDNFDINGLEFGYGFEDILIKIISODPIIKVE
OETPVKITSNSIOEAPWGLHRYHREPKPYGQDLFVLYDAGKGVTSVLDGIDT
EHEDFGRRAEGRATIPANDASDLNNGHTHCAGLTSGKHGVAKNTIYAVKYLRSNG
EGTIVSDVIGIEYVTKHIESKRNKEFGSTANLSLGSKSLAMEMANAADSCV
HFLAIAAGNEDEDALSSPAGAEKSIITGASTFSDDRAFFSNWGCVDVFAPGINIMST
YIGSRNATLSLGSFASNASPVAGILSLFISLOPAPDSEFPNDAPSPELKEKYLKFSI
OGVLGQIDGDDPNKLIYNGGKKLIDGW"
11661..12625
/gene="UNB254"
11661..12625
/gene="UNB254"
note="ORF UNB254, unidentified function"
/codon_start=1
/product="hypothetical protein UNB254"
/protein_id="AAC49483.1"
/db_xref="PID:g1150999"

/db_xref="GI:1150939"
 /translation="MRKRAKSYRKQILVYSHTEFREPIYOVLDNQVLVLECNSEN
 LPSGLKRLQADKAVYITQCCIALYETRNQALINLAQFEHRCNHSFKDPKSAEC
 LESVNIIGANKHRYVVAQODIDRLKRLTVPGVYLIHLTSVMYELSTASASK
 ITEBQKLYGLNDPNIEKLEQSDGSGKESITKRRKLKAPNPLSVKRRKKNVNSPD
 EVKQEDTSKEKKRRKRRKSNTPVNSNGTTAAQ"
 complement(12778..14136)
 /gene="UNE452"
 complement(12778..14136)
 /note="ORF UNE452, unidentified function"
 /codon_start=1
 /product="hypothetical protein UNE452"
 /protein_id="AAC49484.1"
 /db_xref="PID:9151000"
 /db_xref="GI:1151001"
 /translation="MGLGLVLDDEEYKRRHQGDSEIVDHSQEKHIQNSRRKVKILS
 FCSANGISOEFKEIDRKTRGHMKRSEVAPPALIEGSKIPAEWIDPSSEVLEI
 KRSIDNETINMOKATNCTLYGCKRIKYDKETDCTYLDLYESTVSNPSYA
 ERSGDLKRRKRRKRLISDSFHONRKQPISTIRPGLFYVSDVYEDTGIRITRAE
 LEKTIYEHGKLLIVILKRHSIGVRLISCTTTECKALIDRGDILHAPMWVLDIA
 YKRLILEPNYCFNVSQKRAVAERKVDCLDSFENDISEKLSLSYKSQLSLPMGE
 LEIDSEVRRPFLFNSRIYVPRKISTEDDIEMKIKLFGKITTDOQSCNLTIIIP
 YTDPLRKDCNNEVEHEKIKEQIKASDITPKIARVAVAPEMVDHSINENQVPEDEPPV
 NY"
 complement(14387..15610)
 /gene="UND407"
 complement(14387..15610)
 /note="ORF UND407, unidentified function"
 /codon_start=1
 /product="hypothetical protein UND407"
 /protein_id="AAC49485.1"
 /db_xref="PID:9151001"
 /db_xref="GI:1151001"
 /translation="MISALDSIPEPQNPAPSPDEFKLCCELLFEVYQINGTAGTK
 SRSFVYELINFEVEMKRTYGNNTYPAVLALPRDRITINIDYVILRICSVILK
 PKNSATEORLMDKORNGKGNLSLYEELAKRAEBSKAITIDNNHILDSISGD
 RFAGRGKSLVSKSEPLACVENSFVLEKTFEDIVLNNRYIGQEHKLLNCMPDAQ
 DYLSVISDLKAVTSKLDPKYRLKDDLSIVGFAFAPQAKRYNLSEKICRLHLD
 FLVEKMDGERIOVHYNNYGESIKFFSRGIDYILGASISGSIQHLRFTSVKE
 CILDEGMYTFDAKRRVILPFGILVKSAREALSFNINNVDFHLYMVDLLXNGTSL
 TPLPHORXOYLNS"
 complement(15947..16888)
 /gene="UND313"
 complement(15947..16888)
 /note="ORF UND313, unidentified function"
 /codon_start=1
 /product="hypothetical protein UND313"
 Query Match 45.3%; Score 24; DB 8; Length 30469;
 Best Local Similarity 68.8%; Pred. No. 7.5;
 Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 Oy 4 AAGCTTAGGATCTTAAGTCACTATCCGCGACGCGATGACATC 51
 Db 17490 AAGATCTGGCATCCACCGCGACCTAGCACCTGACTGCATGACATC 17337

AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 SOURCE
 Boore, J.L. and Brown, W.M.
 Complete DNA sequence of the mitochondrial genome of the black
 chiton, *Katharina tunicata*
 Genetics 138 (2), 423-443 (1994)
 95129806
 2 (bases 1 to 15532)
 Boore, J.L.
 Direct Submission
 Submitted (17-MAY-1994) Jeffrey L. Boore, Cell Biology and
 Neuroanatomy, University of Minnesota, 321 Church St. SE,
 Minneapolis, MN 55455, USA
 Location/Qualifiers
 1. 15532
 /organism="Katharina tunicata"
 /mitochondrion
 /db_xref="taxon:34587"
 1. 1548
 /note="followed by 'T' as an abbreviated stop codon"
 /codon_start=1
 /translation="MGLGLVLDDEEYKRRHQGDSEIVDHSQEKHIQNSRRKVKILS
 FCSANGISOEFKEIDRKTRGHMKRSEVAPPALIEGSKIPAEWIDPSSEVLEI
 KRSIDNETINMOKATNCTLYGCKRIKYDKETDCTYLDLYESTVSNPSYA
 ERSGDLKRRKRRKRLISDSFHONRKQPISTIRPGLFYVSDVYEDTGIRITRAE
 LEKTIYEHGKLLIVILKRHSIGVRLISCTTTECKALIDRGDILHAPMWVLDIA
 YKRLILEPNYCFNVSQKRAVAERKVDCLDSFENDISEKLSLSYKSQLSLPMGE
 LEIDSEVRRPFLFNSRIYVPRKISTEDDIEMKIKLFGKITTDOQSCNLTIIIP
 YTDPLRKDCNNEVEHEKIKEQIKASDITPKIARVAVAPEMVDHSINENQVPEDEPPV
 NY"
 /db_xref="GI:1150939"
 /translation="MRKRAKSYRKQILVYSHTEFREPIYOVLDNQVLVLECNSEN
 LPSGLKRLQADKAVYITQCCIALYETRNQALINLAQFEHRCNHSFKDPKSAEC
 LESVNIIGANKHRYVVAQODIDRLKRLTVPGVYLIHLTSVMYELSTASASK
 ITEBQKLYGLNDPNIEKLEQSDGSGKESITKRRKLKAPNPLSVKRRKKNVNSPD
 EVKQEDTSKEKKRRKRRKSNTPVNSNGTTAAQ"
 complement(12778..14136)
 /gene="UNE452"
 complement(12778..14136)
 /note="ORF UNE452, unidentified function"
 /codon_start=1
 /product="hypothetical protein UNE452"
 /protein_id="AAC49484.1"
 /db_xref="PID:9151000"
 /db_xref="GI:1151001"
 /translation="MGLGLVLDDEEYKRRHQGDSEIVDHSQEKHIQNSRRKVKILS
 FCSANGISOEFKEIDRKTRGHMKRSEVAPPALIEGSKIPAEWIDPSSEVLEI
 KRSIDNETINMOKATNCTLYGCKRIKYDKETDCTYLDLYESTVSNPSYA
 ERSGDLKRRKRRKRLISDSFHONRKQPISTIRPGLFYVSDVYEDTGIRITRAE
 LEKTIYEHGKLLIVILKRHSIGVRLISCTTTECKALIDRGDILHAPMWVLDIA
 YKRLILEPNYCFNVSQKRAVAERKVDCLDSFENDISEKLSLSYKSQLSLPMGE
 LEIDSEVRRPFLFNSRIYVPRKISTEDDIEMKIKLFGKITTDOQSCNLTIIIP
 YTDPLRKDCNNEVEHEKIKEQIKASDITPKIARVAVAPEMVDHSINENQVPEDEPPV
 NY"
 complement(14387..15610)
 /gene="UND407"
 complement(14387..15610)
 /note="ORF UND407, unidentified function"
 /codon_start=1
 /product="hypothetical protein UND407"
 /protein_id="AAC49485.1"
 /db_xref="PID:9151001"
 /db_xref="GI:1151001"
 /translation="MISALDSIPEPQNPAPSPDEFKLCCELLFEVYQINGTAGTK
 SRSFVYELINFEVEMKRTYGNNTYPAVLALPRDRITINIDYVILRICSVILK
 PKNSATEORLMDKORNGKGNLSLYEELAKRAEBSKAITIDNNHILDSISGD
 RFAGRGKSLVSKSEPLACVENSFVLEKTFEDIVLNNRYIGQEHKLLNCMPDAQ
 DYLSVISDLKAVTSKLDPKYRLKDDLSIVGFAFAPQAKRYNLSEKICRLHLD
 FLVEKMDGERIOVHYNNYGESIKFFSRGIDYILGASISGSIQHLRFTSVKE
 CILDEGMYTFDAKRRVILPFGILVKSAREALSFNINNVDFHLYMVDLLXNGTSL
 TPLPHORXOYLNS"
 complement(15947..16888)
 /gene="UND313"
 complement(15947..16888)
 /note="ORF UND313, unidentified function"
 /codon_start=1
 /product="hypothetical protein UND313"
 Query Match 45.3%; Score 24; DB 8; Length 30469;
 Best Local Similarity 68.8%; Pred. No. 7.5;
 Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 Oy 4 AAGCTTAGGATCTTAAGTCACTATCCGCGACGCGATGACATC 51
 Db 17490 AAGATCTGGCATCCACCGCGACCTAGCACCTGACTGCATGACATC 17337

LLVLSYSDHPY"
complement(3758..3822)
/note="codon recognized: Trc"
/product="tRNA-Phe"
/anticodon="(pos:3789..3791,aa:Phe)
complement(3821..5536)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 5"
/protein_id="AAC48366.1"
/db_xref="pid:9557278"
/db_xref="gi:557278"
/translation="MKAMFLISQATVLLISLSLSTAMTTLTKKILLEFCFYN
FNSVITLPLIMTALFSAIVCTISAVMLETNYSODIFKRFIVWLFLVLS
NLIFIPSLATLLQMDGLGLVSCFVLYONYKSLGAGLMTAFNRRIGDALIAS
VTFSGHFNALFMSTPMITLISMLLAGMKRSQALPESWLPAAMAAPVSAVH
SSITVAGVFLIIQYPLNSHLSTSLMTISMTWAGISANFEDDKIKIALST
LSQGVMLLSVSLGFLFALFPLFHAMEKALFECAGNIHSHNNODIRKSHMT
OMPSSTCFNANLALCGPPMAGFYSKDVITEMMFNONFLISLNFALITLASY
SARSLFTFWSKMQSSVSTSDENSIAPITMLAAGATSGAVFSWLILPSTPL
ISMNKLMAIFATLGAATYKLFLEKISHSIINFEFNMFMINLSNNMTSKSSL
YTGITFESLDGMEITIGEGILNKIKMTLTITNOKNSNLENFISLTLMLFISV
TMWFL"
complement(5537..5603)
/note="codon recognized: CAC"
/product="tRNA-His"
/anticodon="(pos:5568..5570,aa:His)
complement(5597..6928)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_except="(pos:complement(6788..6790),aa:Trp)
/transl_table=5
/product="NADH dehydrogenase subunit 4"
/protein_id="AAC48367.1"
/db_xref="pid:9335324"
/db_xref="gi:335324"
/translation="MSRSTVTKNSKMKWTKMSISHPMPTLNFSPNIRKIN
TEWSDNFSVPLMLSCWITGLIMSSYKILKNLVSFLLNVLNFIIMVTK
SLFSYIFFEASLIPITLLILMAGYQPELQGMVMTYTLIGALPFLINFIESH
AHLNLMMSLIPIMPQAMISFWMLFILLVFLKPLISFLMLPKRAVEAPVAGSMI
LAALLKLGAGINTINNINIKRYNFVNLFFITLGVGVSSLCICQSDMKALIAV
SSVGMGIMAGLISSPERGLKMLIMTANHLSGSGFSSNMNMYEKSQSGSLITK
GPISLVPTFSKSFELMCSTINMAAPSTINLSAGMLISVSLSKFLMLLAMAFAVA
VYTLFTNTQHGKISSVNPNNPKPLYNYVFLHWPLOLLILSTTIC"
complement(6940..7242)
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 4L"
/protein_id="AAC48368.1"
/db_xref="pid:9557280"
/db_xref="gi:557280"
/translation="MMNFNPLFTLGALSSISALMTICQKRLNTLLALEILWVSI
FLVLSINIMSEGLMFVLLTLAASEISGLAVILIRAHGDIYSSISIHNC"
7247..7312
/note="codon recognized: ACA"
/product="tRNA-Thr"
/anticodon="(pos:7275..7277,aa:Thr)
complement(7312..7378)
/note="codon recognized: TCA"
/product="tRNA-Ser"
/anticodon="(pos:7348..7350,aa:Ser)
complement(7381..8520)
/codon_start=1
/transl_except="(pos:complement(8176..8178),aa:Trp)
/transl_table=5
/product="cytochrome b"
/protein_id="AAC48371.1"
/db_xref="pid:9335325"
/db_xref="gi:335325"
/translation="MKPIRSHPVFKIINGSLIDLPSPNLSIMWFGSLGLCII
OITGLFLAMHYATNIELAFSSVAHICGVNNGMJJRAHANAGSLFLFCYFHIHGRG
MYGSGFVNHTMWGVILLFLMTGTAFIGTYVPRQCMGWCATVITNLSAIPYIGTT

MYWIMGFAVDNATLRFSPHYLAPITIAASVYHILFIHSGSNPLGINSSEK
VTISFTLKDVGFVAVNITLLVLEFDPLHAPBDFIPANPLVTFTHQPEWTEL
FAVALRSIIPKRLGVGLMSIVILEFVPLINAGKEPTISFPVTOQLVMSLVSEL
LITWAGAPVEDPDLIGQILTFVYFLFLFIMIPILTKKMDWTINE"
complement(8513..9019)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 6"
/protein_id="AAC48369.1"
/db_xref="pid:9557281"
/db_xref="gi:557281"
/translation="MTFLTMSLFSQAFSIPULPOTPSGLLFLMSLFSILFEC
SSVSEFLFLYIGGLVMAFYATLMPNLMFKNLIQVFPVSTVWMLMLTUSD
IGTLLEQTKMFLMNSFENHGISLSPFMLIIVSALILFVILICVAKICFSNG
PLRPKYVA"
9024..9089
/note="codon recognized: CCA"
/product="tRNA-Pro"
/anticodon="(pos:9054..9056,aa:Pro)
complement(9082..10032)
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 1"
/protein_id="AAC48370.1"
/db_xref="pid:9557282"
/db_xref="gi:557282"
/translation="MFLSFMVAVIATILAAVAFETLLERGLGYIQRRKPNKVS
IGLAPADALKFLKKNLNYNSKRLPYAPVLSLISCLMSLWSDPLWYFN
GVLPETLSALSYGCTMAGASNSKVALGATAPVACTIEYVTLFVILSSAALS
SLNLOISSHOLLYNSLITMLPPLMLMIVCAETNRPAPPBABESLVSQFNIEG
SGSEFLLEAGVASLIFMSMTITLFLGASGWIITNLEWAKVFLAFVITWKCIF
PLRLYDLMLKLTWLSFLPSLIMTLMNLSARFMTLLS"
complement(10033..10101)
/note="codon recognized: TRA"
/product="tRNA-Leu"
/anticodon="(pos:10068..10070,aa:Leu)
complement(10101..10165)
/note="codon recognized: CTA"
/product="tRNA-Leu"
/anticodon="(pos:10132..10134,aa:Leu)
complement(10166..11440)
/note="The exact endpoints are uncertain"
Query Match 44.9%; Score 23.8; DB 37; Length 15532;
Best Local Similarity 72.1%; Pred. No. 9.2;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 TTTAAGCTTAGGCACTTCTAAGTCACTATCCGACCTCGAG 43
DB 10306 TTTAAGCTTCTGCACTTAAGTACTTAAACCAACATCGAG 10348
RESULT 12
CH068040 9829 bp DNA PLN 16-SEP-1996
LOCUS Cochliobolus heterostrophus polyketide synthase (PKS1) gene,
DEFINITION complete cds.
ACCESSION U68040
NID g1546071
VERSION 068040.1 GI:1546071
KEYWORDS
SOURCE Cochliobolus heterostrophus.
ORGANISM Cochliobolus heterostrophus
Eukaryota; Fungi; Ascomycota; Euascomycetes; Loculoascomycetes;
Pileosporales; Pileosporaceae; Cochliobolus.
REFERENCE
1 (bases 1 to 9829)
Yang,G., Rose,M.R., Turgeon,B.G. and Yoder,O.C.
TITLE Cochliobolus heterostrophus PKS1
JOURNAL Plant Cell (1996) In press
2 (bases 1 to 9829)
Yang,G., Rose,M.R., Turgeon,B.G. and Yoder,O.C.
AUTHORS
TITLE Direct Submission


```

Query Match          43.0%; Score 22.8; DB 12; Length 213;
Best Local Similarity 66.0%; Pred. No. 26;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      4  AAGCTTAGGCAATTCCTAAAGTCACATCATCCCGACGAGTCGACATCGA 53
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      121 AAACTCAGGAATCCTCCAGACAGTCCTCTTTATCCGAGTCACATAGA 72

RESULT 15
AF155922/c
LOCUS      AF155922      213 bp      DNA      ROD      16-JUN-1999
DEFINITION Sigmodon hispidus MHC class II antigen (Sih1-DQA) gene, Sih1-DQA*09
            allele, partial cds.
ACCESSION  AF155922
NID         95070687
VERSION     AF155922.1  GI:5070687
KEYWORDS
SOURCE      Sigmodon hispidus.
            Sigmodon hispidus.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognath; Muridae; Sigmodontinae; Sigmodon.
REFERENCE   1 (bases 1 to 213)
AUTHORS    Piau,R.S., Van Den Bussche,R.A., McBee,K. and Lochmiller,R.L.
TITLE      Allelic diversity at the Mhc-DQA locus in cotton rats (Sigmodon
            muridae) and a comparison of DQA sequences within the family
            Muridae (Mammalia: Rodentia)
JOURNAL    Immunogenetics (1999) In press
REFERENCE   2 (bases 1 to 213)
AUTHORS    Piau,R.S., Van Den Bussche,R.A., McBee,K. and Lochmiller,R.L.
TITLE      Direct Submission
JOURNAL    Submitted (02-JUN-1999) Zoology, Oklahoma State University, LSW
            430, Stillwater, OK 74078, USA

FEATURES
     source          location/Qualifiers
     gene            1..213
                        /organism="Sigmodon hispidus"
                        /db_xref="taxon:42415"
     cds             1..213
                        /gene="Sih1-DQA"
                        /number=2
     gene            <1..>213
                        /gene="Sih1-DQA"
                        /allele="Sih1-DQA*09"
                        <1..>213
                        /gene="Sih1-DQA"
                        /product="MHC class II antigen"
                        <1..>213
                        /gene="Sih1-DQA"
                        /codon_start=2
                        /product="MHC class II antigen"
                        /protein_id="AA039252.1"
                        /db_xref="PID:95070688"
                        /db_xref="GI:5070688"
                        /translation="GMAIYSHKPNQGYIFEPDDEKFYVDSDKETVWHIPEFGELT
                        SFDPOGLDIAIVKHLIELTKRSNS"
BASE COUNT      66 a      42 c      52 g      53 t
ORIGIN

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:21:48 ; Search time 133.08 Seconds
(without alignments)
99.641 Million cell updates/sec

Title: US-09-325-095-1718.SEO

Perfect score: 53
Sequence: 1 TTTAAGCTTAGGCATTCTTAA.....CCGACTCGATCGACATCGCA 53

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being pinned,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	66.0	35	1	Q59016
2	35	66.0	35	1	Q57022
3	26	49.1	3207	1	Q51155
4	26	49.1	3498	1	Q57012
5	24.4	46.0	3412	1	Q51156
6	24.4	45.3	1041	1	X15656
7	21.2	40.0	4130	1	V62749
8	20.8	39.2	11531	1	Q54222
9	20.8	39.2	6210	1	Q63815
10	20.6	38.9	53585	1	X20251
11	20.6	38.9	433	1	V64625
12	20.6	38.9	1906	1	V64625
13	20.2	38.1	2204	1	Q11675
14	20.2	38.1	2285	1	Q45935
15	20.2	38.1	2204	1	Q89532
16	19.8	37.4	225	1	Q31806
17	19.8	37.4	468	1	Q31811
18	19.8	37.4	314	1	Q77252
19	19.8	37.4	1628	1	Q74628
20	19.8	37.4	10669	1	V52190
21	19.6	37.0	1755	1	N91166
22	19.6	37.0	4648	1	Q12756
23	19.6	37.0	4648	1	Q12770
24	19.6	37.0	1196	1	T62680
25	19.6	37.0	1196	1	T62681
26	19.6	37.0	2089	1	V06591
27	19.6	37.0	11307	1	X12986
28	19.6	37.0	110000	1	X20248.01
29	19.4	36.6	7499	1	N91312
30	19.4	36.6	7824	1	Q74052
31	19.4	36.6	14690	1	X22303
32	19.4	36.6	10620	1	X22303
33	19.2	36.2	1423	1	O67209
34	19.2	36.2	1423	1	O67210
35	19.2	36.2	1421	1	O71698
36	19.2	36.2	133894	1	T13635
37	19.2	36.2	545	1	V53327
38	19.2	36.2	400	1	V78144
39	19.2	36.2	5137	1	V78616
40	19.2	36.2	5137	1	V74528
41	19.2	36.2	13856	1	V74342
42	19.2	35.8	1714	1	N90718
43	19.2	35.8	1847	1	N90719

ALIGNMENTS

C 44	19	35.8	363	1	Q10877
C 45	19	35.8	5534	1	Q13574
RESULT 1					
ID	Q59016	standard; DNA; 35 BP.			
AC	Q59016:				
DT	12-APR-1994	(first entry)			
DE	RACE PCR primer 2280.				
KW	phosphoinositide kinase; PI; p85 subunit; screening; agonist;				
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;				
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;				
KW	ss; amplification; p110.				
OS	Synthetic.				
PN	W09321328-A.				
PD	28-OCT-1993.				
PF	13-APR-1993; G00761.				
PR	13-APR-1992; GB-008135.				
PA	(LUDW-) LUDWIG INST CANCER RES.				
PI	Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;				
PI	ParkerPJ, Volinia S, Waterfield MD;				
DR	WPI; 93-351738/44.				
PT	Recombinant polypeptide(s) - with phosphoinositide-3 kinase				
PS	activity, useful for controlling cell proliferation				
CC	Example 1: Page 39; 146pp: English.				
CC	An SGBAF-1 cell line was established by transfection of bovine				
CC	adrenal cortex zona fasciculata cells with pSV3neo. Total RNA was				
CC	isolated from the SGBAF-1 and a cDNA library constructed. RACE PCR				
CC	was performed by synthesising first strand cDNA from random hexamers				
CC	on SGBAF-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-				
CC	dr as primers. Products were fractionated using an agarose gel. The				
CC	cDNA was isolated from the gel and subjected to PCR using oligo 2280 and				
CC	adaptor as primers. A product of 350 bp was further sequenced.				
CC	See also Q51155-6, Q59012-23 and Q57522-3.				
CC	Sequence 35 BP; 10 A; 9 C; 4 G; 12 T;				
Query Match					
Best Local Similarity 100.0%; Pred. No. 3.8e-06;					
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	TTTAAGCTTAGGCATTCTTAAGTCACTATCATCC 35			
DB	1	TTTAAGCTTAGGCATTCTTAAGTCACTATCATCC 35			
RESULT 2					
ID	Q57022	standard; DNA; 35 BP.			
AC	Q57022:				
DT	31-AUG-1994	(first entry)			
DE	Ptdins 3-kinase oligo 2280				
KW	110 kD catalytic subunit; phosphatidylinositol 3-kinase;				
KW	transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;				
KW	ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;				
KW	blood vessel plaques; ss.				
OS	Synthetic.				
PN	W09403609-A.				
PD	17-FEB-1994.				
PF	05-AUG-1993; G01651.				
PR	05-AUG-1992; GB-016654.				
PA	(IMCR) IMPERIAL CANCER RES TECHNOLOGY.				
PI	Goode NT, Nurse PM, Parker PJ, Waterfield MD;				
PI	WPI; 94-065697/08.				
DR	Eukaryotic cells transfected with mammalian phospholipid or				
PT	protein kinase DNA - useful in assays for compounds involved in				
PT	cell growth regulation and for treating cancers				
PS	Example 1: Page 15; 71pp: English.				
CC	The sequences given in Q57020-23 are primers which were used in the				

CC Isolation of the phosphatidylinositol (PtdIns) 3-kinase cDNA. The
 CC amplified sequence was placed under the regulatory control of the nmt
 CC promoter. These sequences were transformed into Schizosaccharomyces
 CC pombe cells in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the PtdIns
 CC plasmids grow as the parental strain. In the absence of thiamine the
 CC nmt promoter functions and the PtdIns is induced. PtdIns activity is
 CC substantially increased under these conditions. Cells containing
 CC constructs such as this, are useful in assays for detecting compounds
 CC involved in cell growth regulation. It is also used as the basis for
 CC detecting compounds for treating cancers and the formation of blood
 CC vessel plaques.
 SO Sequence 35 BP; 10 A; 9 C; 4 G; 12 T;

Query Match 66.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 3.8e-06;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTTAGCTTAGCATTTAAAGTCATCATCC 35
 DB 1 TTTAGCTTAGCATTTAAAGTCATCATCC 35

RESULT 3
 ID 051155 standard; cDNA; 3207 BP.
 AC 051155;
 DT 12-APR-1994 (first entry)
 DE p110 cDNA.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KM antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KM platelets; neurorphil activity; 3-phosphorylated phosphoinositides;
 KM ss.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /note= "PI3- kinase p110"

PN MO9321328-A.
 PD 28-OCT-1993.
 PR 13-APR-1993; G00761.
 PA (LUDWIG) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinina S, Waterfield MD;
 PI P-PSDB; R43341.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Claim 7; Fig 9; 146pp; English.
 CC Southern blot analysis was performed using a bovine cDNA probe contg.
 CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
 CC from a cDNA library constructed from mRNA isolated from the human
 CC cell line K562. Positive clones were sequenced to give the human
 CC PI3 kinase p110 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p110 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to
 CC provide a protein with PI3 kinase activity, and is useful for
 CC screening for (ant)agonists of PI3 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC prophylaxis or therapy. Platelet or neutrophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051156, 059012-23 and 057522-3.
 SO Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

DB 110 AGGCATCTAAAGTCATCATCC 85

RESULT 4
 ID 057012/c
 AC 057012 standard; cDNA to mRNA; 3498 BP.
 DT 31-AUG-1994 (first entry)
 DE PtdIns 3-kinase 110 kD catalytic subunit cDNA.
 KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
 KM transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KM PtdIns 3-kinase; assay; detection; cell growth; regulation; cancer;
 KM blood vessel plaques; ss.
 OS Bos taurus.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /product= p110

PN MO9403609-A.
 PD 17-FEB-1994.
 PR 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 PI WPI: 94-065697/08.
 DR P-PSDB; R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure: Fig 1; 71pp; English.
 CC This sequence encodes the 110 kD catalytic subunit of the phosphatidylin
 CC inositol (PtdIns) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nmt promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the PtdIns
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nmt promoter functions and the PtdIns 3-kinase
 CC catalytic subunit is induced. PtdIns activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SO Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 49.1%; Score 26; DB 1; Length 3498;
 Best Local Similarity 100.0%; Pred. No. 0.076;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AGGCATCTAAAGTCATCATCC 35
 DB 110 AGGCATCTAAAGTCATCATCC 85

RESULT 5
 ID 051156/c
 AC 051156 standard; cDNA; 3412 BP.
 DT 12-APR-1994 (first entry)
 DE Human p110 cDNA.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KM antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KM platelets; neurorphil activity; 3-phosphorylated phosphoinositides;
 KM ds.
 OS Human.

FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /note= "PI3- kinase p110"
 PN MO9321328-A.
 PD 28-OCT-1993.
 PR 13-APR-1993; G00761.

Query Match 49.1%; Score 26; DB 1; Length 3207;
 Best Local Similarity 100.0%; Pred. No. 0.075;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AGGCATCTAAAGTCATCATCC 35
 DB 110 AGGCATCTAAAGTCATCATCC 85

PR 13-APR-1992; GB-008135.
 PA (LUDWIG) INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 UK P-PSDB: RA3342.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity; useful for controlling cell proliferation
 PS Claim 7; Fig 16; 145pp; English.
 CC Southern blot analysis was performed using a bovine cDNA probe contg.
 CC a fragment of a p13-kinase-encoding sequence and human cDNA isolated
 CC from a cDNA library constructed from mRNA isolated from the human
 CC cell line K562. Positive clones were sequenced to give the human
 CC p13 kinase p110 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p110 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to
 CC provide a protein with p13 kinase activity, and is useful for
 CC screening for (ant)agonists of p13 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC for prophylaxis or therapy. Platelet or neutrophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051155 and 057522-3.
 SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 46.0%; Score 24.4; DB 1; Length 3412;
 Best Local Similarity 96.2%; Pred. No. 0.36;
 Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 AGGCATTCCTAAGTCACTATCATCC 35
 DB 110 AGGCATTCCTAAGTCACTATCATTC 85

RESULT 6
 X15656/C
 ID X15656 standard; DNA: 1041 BP.
 AC X15656;
 DT 07-MAY-1999 (first entry)
 DE Protein phosphatase gene coding sequence of *Saccharomyces cerevisiae*.
 KW Protein phosphatase gene; growth; fermentation activity;
 KW dough production; yeast; ds.
 OS *Saccharomyces cerevisiae*.
 PN J11042090-A.
 PD 16-FEB-1999.
 PF 29-JUL-1997; 203652.
 PR 29-JUL-1997; JP-203652.
 RA (KANF) KANEKA CORP.
 RA (SHOS) SHOWA SANGYO CO.
 DR WPI: 99-197822/17.
 PT New yeast of controlled activation at low temperatures - useful for
 PT improving the quality of dough
 PS Claim 3; Page 19-20; 41pp; Japanese.
 CC The present sequence represents a protein phosphatase gene sequence
 CC of *Saccharomyces cerevisiae*. The specification describes new
 CC *S. cerevisiae* in which the growth and/or the fermentation activity
 CC is controlled at least in the range of 0-20 degrees Celsius. These
 CC yeast are prepared by deleting the function of at least one protein
 CC phosphatase gene. The yeast is useful in the production of dough.
 SQ Sequence 1041 BP; 553 A; 184 C; 243 G; 261 T;

Query Match 45.3%; Score 24; DB 1; Length 1041;
 Best Local Similarity 68.8%; Pred. No. 0.39;
 Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 4 AAGCTTAGCATTCCTAAGTCACTATCATCCGACTCGATCGACATC 51
 DB 756 AAGATCTGGCAATCCACCGCACTAGCACCTGATTCGATCGACATC 709

RESULT 7

V62749/C
 ID V62749 standard; CDNA: 4130 BP.
 AC V62749;
 DT 15-FEB-1999 (first entry)
 DE Human secreted protein clone er418.5 CDNA.
 KW Secreted protein; human; er418.5; ds.
 OS Homo sapiens.
 FT Key
 FT CDS
 FT Location/Qualifiers
 FT /tag- a
 PN W09846757-A2.
 PD 22-OCT-1998.
 PR 12-APR-1998; 007999.
 PR 13-APR-1998; US-059487.
 PR 15-APR-1997; US-843374.
 PA (GENW) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, LaValle ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI: 98-568731/48.
 DR P-PSDB: W74723.
 PT New polynucleotide(s) encoding secreted human proteins - are derived
 PT from, e.g. human foetal brain or foetal kidney cDNA libraries,
 PT potentially useful as, e.g. vaccines or thrombolytic agents
 PS Claim 19; Page 76-78; 120pp; English.
 CC Full-length cDNA clone er418.5 includes an open reading frame
 CC encoding a human secreted protein (see W74723). It was isolated
 CC from a human foetal brain cDNA library using methods which are
 CC selective for cDNAs encoding secreted proteins, or was identified
 CC as encoding a secreted or transmembrane protein on the basis of
 CC computer analysis of the amino acid sequence of the encoded protein.
 CC It shows some similarity to database sequences. The invention
 CC provides polynucleotides (see V62746-55) from human foetal brain,
 CC adult testis, adult brain, adult kidney and foetal kidney (all
 CC deposited as composite clone ATCC 98404), which encode human
 CC secreted proteins (see W74720-29). The polynucleotides and
 CC proteins are predicted to have biological activities which would
 CC make them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data are
 CC given. Suggested activities include nutritional activity, immune
 CC stimulating (e.g. as vaccines) or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activator/inhibitor activity, chemotactic/chemokinetic activity,
 CC haemostatic and thrombolytic activity, receptor/ligand activity,
 CC antiinflammatory activity, cadherin/tumour invasion suppressor
 CC activity, and tumour inhibition activity. The polynucleotides are
 CC also stated to be useful for gene therapy, and can be used in
 CC recombinant production of the polypeptides.
 SQ Sequence 4130 BP; 1113 A; 1051 C; 1024 G; 936 T;

Query Match 40.0%; Score 21.2; DB 1; Length 4130;
 Best Local Similarity 64.0%; Pred. No. 8.4;
 Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 TTAAGCTTAGCATTCCTAAGTCACTATCATCCGACTCGATCGACATC 51
 DB 794 TTAATCTTAGCATTCCTAAGGAGACCTTCCTCTGCTGACGATC 745

RESULT 8
 054222/C
 ID 054222 standard; DNA: 11531 BP.
 AC 054222;
 DT 27-JUN-1994 (first entry)
 DE BSSU/CEL Gene.
 KW BSSU/CEL Gene.
 KW BSSU, biologically functional bile salt stimulated lipase; CEL;
 KW carboxyl ester lipase; hybridisation; milk protein; transgenic;
 KW infant milk substitute; oral; specific; lipolysis; vitamin; bile;
 KW lipid malabsorption; cystic fibrosis; chronic pancreatitis;
 OS digestion; ss.
 OS Homo sapiens.
 FH Key
 FH CDS
 FH Location/Qualifiers
 FH cds 1653..11397

PN M09858943-A1.
PD 30-DEC-1998. U12764.
PE 18-JUN-1998; US-057483.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Iathigra R, Smith HO,
PI White OR;
DR WPI: 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1: Page 801-831; 1128pp; English.
CC X02048 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (hb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
SQ Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T;

Query Match 39.2%; Score 20.8; DB 1; Length 53585;
Best Local Similarity 70.0%; Pred. No. 24;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 10 AGGATTCTAAAGTCACTATCATCCGAGTCGATCGACA 49
44697 AAGCTTCAGAGAACCATCATCCACAGATGACACA 44658
|||||

RESULT 11
ID V31302 standard; DNA; 433 BP.
AC V31302;
NM 01-OCT-1998 (first entry)
DE E. coli J96 pathogenicity island contig #116.
KW PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pher;
KW PAI V; pher; vaccine; protective immune response; ds.
OS Escherichia coli.
FM WC9622575-A2.
PD 28-MAY-1998.
PE 21-NOV-1997; U21347.
PR 14-OCT-1997; US-061953.
PR 22-NOV-1996; US-031626.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYMI-) UNIV WISCONSIN.
PI Choi GH, Dillon PJ, Welch RA;
PI WPI: 98-312461/27.
DR New isolated uropathogenic E. coli nucleotide sequences - used to
DR develop products for the detection of pathogenic E. coli and to
PT elicit an immune response to pathogenic E. coli
PS Claim 21: Page 226; 250pp; English.
CC This sequence represents a E. coli strain J96 contig containing
CC pathogenicity island (PAI) sequences, and represents a nucleic acid
CC molecule of the invention. PAIs are large fragments of DNA which comprise
CC pathogenicity determinants. The sequences of the invention are taken from
CC PAI IV and PAI V. PAI IV is located at approximately 64 min (near pher)
CC on the E. coli chromosome and is greater than 170 kb. PAI V is located at
CC approximately 94 min (at pher) on the E. coli chromosome and is
CC approximately 160 kb in size. Antibodies specific to the proteins encoded
CC by the PAI open reading frames of the invention can be used in kits to
CC detect uropathogenic E. coli. The proteins are used in vaccines to elicit
CC a protective immune response in an animal to the uropathogenic E. coli
CC strain J96.
SQ Sequence 433 BP; 129 A; 94 C; 64 G; 145 T;

Query Match 38.9%; Score 20.6; DB 1; Length 433;
Best Local Similarity 62.7%; Pred. No. 8.4;
Matches 32; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 TTAAAGTCTAGGCTTAAGTCACTATCATCCGAGTCGATCGACATC 51
289 TTCAGCTTATACATGATGATGATCATCTTCTGCTGAGGCCACATC 339
|||||

RESULT 12
ID V64625 standard; cDNA; 1906 BP.
AC V64625;
NM 02-MAR-1999 (first entry)
DE Tomato fructokinase (Fk1) cDNA.
KW Tomato; Polato; fructokinase; regulatory sequence; ripening; fruit;
KW carbohydrate metabolism; transgenic plant; expression cassette;
KW Lycopersicon esculentum; ss.
OS Lycopersicon esculentum.
FM Key Location/Qualifiers
FT CDS 271..1314
FT /*tag- a
FT /*note- "Fructokinase (fk1) coding region"

PN M09845412-A1.
PD 15-OCT-1998.
PE 10-MAR-1998; U04649.
PR 05-APR-1997; US-826611.
PA (REGC) UNIV CALIFORNIA.
PI Bennett AB, Kanayama Y;
PI WPI: 98-583193/49.
DR P-PSDB; W81785.
DR New fructokinase gene from tomato - useful for, e.g. inhibiting
PT fructokinase activity for increasing sweetness or stimulating starch
PT biosynthesis in plants
PS Claim 1: Page 28-32; 51pp; English.
CC The introduction of an expression cassette containing an exogenous Fk1
CC (fructokinase) sequence, is used to modify sweetness and promote starch
CC biosynthesis in transgenic plants, particularly tomato. Expression of
CC Fk1 in sense or antisense orientation inhibits conversion of fructose
CC to its 6-phosphate, increasing sweetness. Overexpression of
CC fructokinase is used to increase starch content in tubers, roots and
CC seeds.
SQ Sequence 1906 BP; 515 A; 362 C; 422 G; 607 T;

Query Match 38.9%; Score 20.6; DB 1; Length 1906;
Best Local Similarity 62.7%; Pred. No. 12;
Matches 32; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 TTAAAGTCTAGGCTTAAGTCACTATCATCCGAGTCGATCGACATC 51
1740 TTAAGTCTAGGCTTAAGTCACTATCATCCGAGTCGATCGACATC 1790
|||||

RESULT 13
ID Q11675/c
AC Q11675; standard; DNA; 2204 BP.
NM 16-JUL-1991 (first entry)
DE H2RIIBP cDNA clone.
KW Major histocompatibility complex; MHC class I regulatory element;
KW nuclear hormone receptors; tumours; cancer; H2RIIBP; ss.
OS synthetic.
FM Key Location/Qualifiers
FT CDS 1..1341
FT /*tag- a
FT /*product- H2RIIBP protein
FT misc_rna 346..549
FT /*tag- b
FT /*label- zinc
FT poly_a_signal 1832..1837
FT /*tag- c

CC	or RXR β at physiological conditions. Sequences encoding 1 isoform	
CC	of mouse RXR β , 3 isoforms of human RXR β , 1 isoform of mouse	
CC	RXR α and 1 isoform of mouse RXR γ are provided (045955-60).	
SO	Sequence 2285 BP; 447 A; 665 C; 681 G; 492 T;	
Query Match	38.1%; Score 20.2; DB 1; Length 2285;	
Best Local Similarity	68.3%; Pred. No. 19;	
Matches 28; Conservative	0; Mismatches 13; Indels 0; Gaps 0	
Oy	6 GCTTAGCATCTTAAGTCACATCATCATCCGACGACGACG 46	
Db	574 GCTTGTCCACGTGTACAGCTTTGTATTATCAGCAGACAGTAG 534	
RESULT 15		
089532/c		
ID	089532 standard; cDNA; 2204 BP.	
AC	089532; (first entry)	
DT	24-OCT-1995	
DE	H-2RIIBP cDNA.	
KM	H-2 region II binding protein; H-2RIIBP; retinoid receptor; ss.	
OS	Not Specified.	
FT	Key	
FT	cds	
FT	Location/Qualifiers	
FT	1..1341	
FT	/*tag- a	
US5403925-A.		
04-APR-1995.		
13-DEC-1989; 450162.		

```

28-SEP-1992; US-952800.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Ozato K;
DR WPI: 95-146856/19.
P-PSDB: R72484.
PT New DNA encoding human H-2 region II binding protein - an
PT Intracellular retinoid receptor mediating effects of steroid(s)
PT etc., also diagnostic kits contg. probes derived from the DNA
PS Disclosure; Column 19-24; 14pp; English.
CC A cDNA library of the Burkitt lymphoma cell line Daudi was screened
CC with a fragment of the murine retinoid X receptor-beta (H-2RIIBP)
CC gene. A positive clone contained an insert (sequence given in
CC Q89531) that encoded human H-2RIIBP (R72483). Further H-2RIIBP
CC sequences of unspecified origin (possibly murine) are given in
CC Q89432 and R72484.
SO Sequence 2204 BP; 436 A; 645 C; 648 G; 475 T;

Query Match 38.1%; Score 20.2; DB 1; Length 2204;
Best Local Similarity 68.3%; Pred. No. 19;
Matches 28; Conservative 0; Mismatches 13; Indels 0; Gaps 0.

QY 6 GCTTAGCGCATCTTAAGTCACACTATCATCCGACGACGATCG 46
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 493 GCTTGCCACGTCGTACAGCTTTGTTATTCACGACGACGAGTAG 453

Search completed: September 13, 1999, 16:21:51
Job time: 4011 sec

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

W. nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:13:52 ; Search time 849.94 Seconds

(without alignments)
123,002 Million cell updates/sec

Title: US-09-325-095-1718.SEQ

Sequence: 1 TTTAAGCTTAGCATTTCTAA.....CCGACTCGATTCGACATCGA 53

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	47.2	417	AA190409	AA190409 zpe9c10.r
2	22.6	42.6	376	AI509866	AI509866 ms30e06.y
3	21.4	40.4	503	AA197781	AA197781 mw03f10.r
4	21.2	40.0	373	DA6325	DA6325 R1C510928A
5	21.2	40.0	275	AA357355	AA357355 EST60068
6	21.2	40.0	430	AA464819	AA464819 z43901.r
7	21	39.6	308	AA513147	AA513147 n129b07.s
8	21	39.6	492	AA987152	AA987152 uc80h09.x
9	20.8	39.2	200	C14053	C14053 C14053 Clon
10	20.8	39.2	404	AA525881	AA525881 n157f06.s
11	20.8	39.2	480	AA585790	AA585790 28439.Lam
12	20.8	39.2	307	AA610665	AA610665 np94a03.s
13	20.8	39.2	360	C61659	C61659 C61659 Yuj1
14	20.8	39.2	407	AA779495	AA779495 ef23b12.s
15	20.8	39.2	484	AA932095	AA932095 o062f01.s
16	20.8	39.2	350	AI025525	AI025525 ov75c10.x
17	20.6	38.9	445	W66698	W66698 me26c08.r1
18	20.6	38.9	594	AA438469	AA438469 LD10349.5
19	20.6	38.9	745	AA438912	AA438912 LD13306.5
20	20.6	38.9	504	AA392739	AA392739 LD11875.5
21	20.6	38.9	334	AA885288	AA885288 a158q04.s
22	20.6	38.9	520	AI104047	AI104047 EST213336
23	20.6	38.9	517	AI104124	AI104124 EST213413
24	20.6	38.9	530	AI105199	AI105199 EST214488
25	20.6	38.9	595	AI172362	AI172362 EST218366
26	20.6	38.9	487	AI232471	AI232471 EST229159
27	20.6	38.9	487	AA819739	AA819739 UT-RN0-2
28	20.6	38.9	555	AI009655	AI009655 EST204106
29	20.6	38.9	599	AI531879	AI531879 SD03164.5
30	20.6	38.9	471	AI722479	AI722479 fc27h07.x
31	20.4	38.5	356	R86883	R86883 Y930a05.r1
32	20.4	38.5	369	R89816	R89816 yp91e02.s1
33	20.4	38.5	366	W34996	W34996 mc34c02.r1
34	20.4	38.5	268	AA224491	AA224491 zt17d03.s
35	20.4	38.5	375	AA350887	AA350887 EST58398
36	20.4	38.5	561	AA518461	AA518461 v106h10.r
37	20.4	38.5	386	C25435	C25435 C25435 R1C5
38	20.4	38.5	664	AA753536	AA753536 96BS0749
39	20.4	38.5	593	AA887564	AA887564 o141c12.s
40	20.4	38.5	817	AI037083	AI037083 ue19b08.y
41	20.4	38.5	609	AU030452	AU030452 AU030452
42	20.4	38.5	545	C97410	C97410 C97410 R1C5
43	20.4	38.5	648	C97445	C97445 C97445 R1C5
44	20.4	38.5	462	C97913	C97913 C97913 R1C5
45	20.4	38.5	394	AU071270	AU071270 AU071270

ALIGNMENTS

RESULT 1
AA190409/c 417 bp mRNA EST 15-JAN-1997
LOCUS zpe9c10.r1 Stratagene Hela cell s3 937216 Homo sapiens cDNA clone
DEFINITION IMAGE:627378 5' similar to SW:P11A BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPNA ISOPORM ;,
mRNA sequence.
ACCESSION AA190409

NID 91779240
 VERSION AAI90409.1 GI:1779240
 KEYWORDS EST.
 SOURCE Human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 417)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B., Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Marls, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rikkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevas, R., Underwood, K., Woldmann, P., Waterston, R., Wilson, R., and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 On May 18, 1995 this sequence version replaced gi:811202.
 COMMENT
 TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 328.
 Location/Qualifiers
 1..417
 /organism="Homo sapiens"
 /db_xref="GDB:5046076"
 /db_xref="taxon:9606"
 /clone="IMAGE:627378"
 /clone.lib="Stratagene Hela cell s3 937216"
 /sex="female"
 /dev_stage="Hela S3 cell line"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: oligo dT. Hela S3 epithelial carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR vector. -5' adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTT 3'."
 BASE COUNT 116 a 98 c 93 g 99 t 11 others
 ORIGIN
 Query Match 47.2%; Score 25; DB 29; Length 417;
 Best Local Similarity 75.6%; Pred. No. 2.7;
 Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 OY 10 AGGCATTCTAACTATCATCCGACCTCGAGTCACAT 50
 |||||
 Db 292 AGGCATTCTAACTATCATCCATTCATTGGTAGAACAAT 252
 RESULT 2
 LOCUS A1509866 376 bp mRNA EST 12-MAR-1999
 DEFINITION MS30606.y1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
 IMAGE:608462 5', mRNA sequence.
 ACCESSION A1509866
 NID 94408771
 VERSION A1509866.1 GI:4408771
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 376)

AUTHORS
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, T., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 On Jun 5, 1998 this sequence version replaced gi:3188558.
 COMMENT
 TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 This read is a RESSEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the correct orientation)
 Seq primer: -40RP from Gibco
 High quality sequence stop: 375.
 Location/Qualifiers
 1..376
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /map="19p12-p13.1"
 /clone="IMAGE:608482"
 /clone.lib="Stratagene mouse skin (#937313)"
 /sex="female"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: oligo dT. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR vector; -5' adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTT 3'."
 BASE COUNT 106 a 97 c 80 g 93 t
 ORIGIN
 Query Match 42.6%; Score 22.6; DB 47; Length 376;
 Best Local Similarity 64.2%; Pred. No. 24;
 Matches 34; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
 OY 1 TTTAGCTTAGCATCTCTAAAGCTATCATCCGACCTCGAGTCACATCGA 53
 |||||
 Db 304 TTAACTCAGGCGAGTTTAAAGCACACCCATCTCCGCCAGATCTACACAGA 356
 RESULT 3
 LOCUS AAI97781 503 bp mRNA EST 12-MAR-1997
 DEFINITION mv03f10.r1 Quaywoodford Beier mouse kidney day 0 Mus musculus cDNA
 clone IMAGE:653995 5', similar to gb:L47650 Mus musculus STAT6 mRNA,
 complete cds (MOUSE);, mRNA sequence.
 ACCESSION AAI97781
 NID 91793423
 VERSION AAI97781.1 GI:1793423
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 503)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R., and Waterston, R.


```

Db          48  AAGATTAAACATTCACAAAGCACAATCTTACTGATACAGTC 89
RESULT      5
LOCUS       AA64819/C
DEFINITION  AA64819 430 bp mRNA EST 10-JUN-1997
             2x3901.r1 Soares_total.fetus_NB2HF8_9w Homo sapiens cDNA clone
             IMAGE:789264 5', mRNA sequence.
ACCESSION   AA64819
NID         92189703
VERSION     AA64819.1 GI:2189703
KEYWORDS    *EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 430)
AUTHORS    Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
             Kuababa,T., Lacey,M., Le,N., Lennon,G., Marra,M., Martin,J.,
             Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theisling,B.,
             White,Y., Wyllie,T., Waterston,R. and Wilson,R.
             Washu-Merck EST Project 1997
             Unpublished (1997)
TITLE       JOURNAL
COMMENT     On May 5, 1995 this sequence version replaced gi:798051.

FEATURES
SOURCE
location/Qualifiers
1..430
/organism="Homo sapiens"
/db_xref="GDB:5966881"
/db_xref="taxon:9606"
/map="6"
/clone="IMAGE:789264"
/clone_11b="Soares_total.fetus_NB2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pTRT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTCACATCTGACGAGTGCAGCGCGCCCTTAATTTTATTTTATTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTR73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatma Bonaldo."

BASE COUNT  130 a      83 c      102 g      115 t
ORIGIN
Query Match      40.0%; Score 21.2; DB 34; Length 430;
Rest Local Similarity 64.0%; Pred. No. 88;
Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY      1  TTTAGCTTAGCATTCTAAAGCTCAATGACCCGAGCTCGATGACAT 50
        |||||
ZL      350  TTTAGTTTGGCAGCAGATGTCTGTGTCATGACTAACTCTTCAT 301

RESULT      7
LOCUS       AA513147/C
DEFINITION  AA513147 308 bp mRNA EST 13-NOV-1997
             n129b07.s1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:971701, mRNA
             sequence.

```

```

ACCESSION      AA513147
NID            92251559
VERSION        AA513147.1  GI:2251559
KEYWORDS
SOURCE
ORGANISM       human.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE          NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT         Tumor Gene Index
                Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLND at:
www-bio11n1.gov/dbfp/image/image.htm

Insert Length: 433      Std Error: 0.00
Seq primer: -40m13 fwd: ET from Amersham.
Location/Qualifiers
1. 308
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:971701"
/clone_lib="NCI-CGAP Ew1"
/tissue_type="Ewing's sarcoma"
/lab_host="DHL0B"
/notice="Vector: PAMPI0; mRNA made from Ewing's sarcoma,
cDNA made by oligo-dT priming. Non-directionally cloned.
Size-selected on agarose gel, average insert size 600 bp
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."

BASE COUNT      74 a      80 c      88 g      66 t
ORIGIN
Query Match      39.6%; Score 21; DB 34; Length 308;
Best Local Similarity 62.3%; Pred. No. 1e+02;
Matches 33; Conservative 0; Mismatches 20; Indels 0; Gaps 0.

Oy 1 TTAACTTAGCATTCTAAAGTCACATCAATCCAGCAGTCGACATCGCA 53
||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 251 TTTCGGCTACAGCGCTTCACAGCTCAGTTCATCATTTCTGCGCGGACGCA 199

RESULT 8
LOCUS      AA987152/c
DEFINITION
AA987152
AA987152
protein mRNA, complete cds (NCUISE);, mRNA sequence.
AA987152
93167915
AA987152.1 GI:3167915
EST.
house mouse.
house musculus
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
Matta, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, D., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Thieling, B., Wylie, J., Lennon, G., Soares, B., Wilson, R. and

```



```

/lab.host="DH10B"
/note="vector: pAMP10; mRNA made from invasive ovarian
tumor, cDNA made by oligo-dT priming. Non-directionally
cloned. Site-selected on agarose gel, average insert size
600 bp. Reference: Rizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT      110 a      90 c      123 g      81 t
ORIGIN

Query Match      39.2%   Score 20.8;   DB 34;   Length 404;
Best Local Similarity 70.0%;   Pred. NO. 1.3e+02;
Matches 28;   Conservative 0;   Mismatches 12;   Indels 0;   Gaps 0;

OY      10  AGGCATTCCTAAGTCACTATCATCCGAGCTCAGTCGACACA 49
          111 111 111 111 111 111 111 111 111 111
Db       64  AGGAATCCCAATTCATTGAGATTCAGACTTGACTTACCA 25

RESULT 11
AA585790/c
LOCUS      28439 Lambda-PRL2 Arabidopsis thaliana cDNA clone 105J12XP 3', MRNA
DEFINITION
ACCESSION AA585790
VERSION    g2393202
KEYWORDS   Arabidopsis
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (bases 1 to 480)
Newman,T., deBrujin,F.J., Green,P., Keegstra,K., Kende,H.,
McIntosh,L., Ohlrogge,J., Ralshel,N., Somerville,S., Thomashow,M.,
Rezeel,E. and Somerville,C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
95148729
On May 5, 1995 this sequence version replaced gi:798282.

REFERENCE
AUTHORS
JOURNAL
MEDLINE
COMMENT

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313rcn@msu.cl.msu.edu
The sequence entry for this EST has been reverse complemented and
is being submitted in the sense orientation.
Seq primer: M13 -21 dye primer.
Location/Qualifiers
1. 480
/organism="Arabidopsis thaliana"
/strain="var columbica"
/db_xref="taxon:3702"
/map="20"
/clone="105J12XP"
/note="Vector: lambda zip-lox; site 1: sal; site 2: Not;
lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
silques). The vector is BRL's lambda zip-lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA. "

```

```

BASE COUNT      146 a       79 c       102 g       142 t       11 others
ORIGIN

Query Match          39.2%; Score 20.8; DB 35; Length 480;
Best Local Similarity 70.0%; Pred. No. 1.3e+02;
Matches    28; Conservative   0; Mismatches 12; Indels     0; Gaps    0;

QY      10 AGGCATTCTAAAGTCACATCATCCCGCAGTGGGACA 49
        ||||||||| | | | | | | | | | | | | | | | |
DB      230 ACTGCTTTAAAGTCTCCATGACACTTACTCGAGTGCAKA 191


RESULT  12
LOCUS   AA610665/c
DEFINITION np44a03.s1 NCI-CGAP-Thy1 Homo sapiens cdna clone IMAGE:1133932 similar to Xr:g1236142 g1236142 MARINER TRANSPOSASE ; contains Alu repetitive element; contains element L1 repetitive element ; mRNA sequence.
ACCESSION AA610665
NID       g2459093
VERSION  g2459093
KEYWORDS  EST.
SOURCE    Human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 307)
AUTHORS  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
TITLE     Unpublished (1997)
JOURNAL  On Sep 12, 1996 this sequence version replaced gi:1394357.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 486-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LMW at: www-bio.liml.gov/bdrp/image/image.html

Insert Length: 543 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 297.
Location/Qualifiers
1..307
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1133932"
/clone_lib="NCI CGAP Thy1"
/tissue_type="thyroid"
/lab_host="DH10B"
/note="Vector: PAMPI0; mRNA made from invasive thyroid tumor, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

```

Db 71 TAAAGTTGGCTTCTCAAGACTCTCATC 40

RESULT 13

LOCUS

661659

DEFINITION

C61659 360 bp mRNA EST 22-SEP-1997
C61659 Yuj1 Kohara unpublished cDNA Caenorhabditis elegans CDNA
clone YK2447 5', mRNA sequence.

ACCESSION

661659

VERSION

92420364

KEYWORDS

EST.

SOURCE

ORGANISM

Caenorhabditis elegans.

REFERENCE

1 (bases 1 to 360)

AUTHORS

Kohara, Y., Mochizuki, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A. and Mishigaki, A.

TITLE

Expression map of the C.elegans genome

JOURNAL

Unpublished (1996)

COMMENT

On May 9, 1995 this sequence version replaced gi:802313.

FEATURES

source

1..360

/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"

/db_xref="taxon:6239"

/clone="YK2447"

/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"

BASE COUNT

104 a 78 c 77 g 101 t

ORIGIN

Query Match

Best Local Similarity 39.2%; Score 20.8; DB 36; Length 360;

Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 14 ATCTAAGTCACTATCATCCGACCTCGACATCGA 53

Db 152 ATCTAAGTCACTATCATCCGACCTCGACATCGA 191

RESULT 14

LOCUS

AA779495

DEFINITION

a123b12.s1 Soares-total_fetus_Nb2HF8_9w Homo sapiens CDNA clone

IMAGE:1032479 3', mRNA sequence.

ACCESSION

AA779495

NID

92838826

VERSION

AA779495.1

KEYWORDS

EST.

SOURCE

ORGANISM

human.

REFERENCE

1 (bases 1 to 407)

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Thesling, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE

Washington University EST Project

JOURNAL

Unpublished (1997)

COMMENT

On Jan 19, 1998 this sequence version replaced gi:2287157.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LINT; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40m13 fwd. Ex from Amersham

High quality sequence stop: 369.

FEATURES

source

1..407

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1032479"

/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"

BASE COUNT

85 a 95 c 123 g 104 t

ORIGIN

Query Match

Best Local Similarity 39.2%; Score 20.8; DB 38; Length 407;

Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 3 TAAAGTCACTATCATCCGACCTCGACATCGA 42

Db 254 TAAAGTCACTATCATCCGACCTCGACATCGA 215

RESULT 15

LOCUS

AA932095

DEFINITION

0062f01.s1 NCI-CGAP Ins Homo sapiens CDNA clone IMAGE:1570777 3'

similar to gp:M35718 FIBROBLAST GROWTH FACTOR RECEPTOR BFR-2

PRECURSOR (HUMAN);, mRNA sequence.

ACCESSION

AA932095

NID

93086408

VERSION

AA932095.1

KEYWORDS

EST.

SOURCE

ORGANISM

human.

REFERENCE

1 (bases 1 to 484)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Tumor Gene Index

JOURNAL

Unpublished (1997).

COMMENT

On Sep 12, 1996 this sequence version replaced gi:1397953.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/BLN at:

www.bio.lnl.gov/db/rp/image/image.html

Insert Length: 794 Std Error: 0.00
Seq primer: -40ml3 fwd. RT from Amersham
High quality sequence stop: 252.

FEATURES

source

1.484
location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1570777"
/clone_lib="NCI CGAP Ins"
/issue_type="carcinoid"

/lab_host="DH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
neuroendocrine lung carcinoid, and was then primed with a
Not I - oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 165 a 104 c 89 g 126 t

ORIGIN

Query Match 39.28; Score 20.8; DB 40; Length 484;
Best Local Similarity 64.6%; Pred. No. 1.3e+02;
Matches 31; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 4 AACCTAGCATCTTAAGTCACATCATCCCGACGAGTCGACATC 51
||| | || |||| | |||| | |||| | |||| | |||| |
DB 116 AAGTTAAGTTATTCACATTCACCACTTCGCCACTTGACCCGCCACC 163

Search completed: September 13, 1999, 16:13:55
Job time: 5454 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:19:36 ; Search time 329.34 Seconds

(without alignments)
386.265 Million cell updates/sec

Title: US-09-325-095-2122.SEQ

Perfect score: 40

Sequence: 1 TGGCTGAATCTTAATGCTGGATTTCATGAACAAATGA 40

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:*
1: gb_bal:*
2: gb_baz:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vl:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pac:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vl:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_bal:*
39: em_baz:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	26	65.0 185092	34 HSJ661G12	AL049797 Homo sapi

C 2	23.8	59.5	28065	11	AC005354	AC005354 Homo sapi
C 3	23.2	58.0	86765	11	AC004472	AC004472 Homo sapi
C 4	23	57.5	110102	7	ATF2812	AL021110 Arabidops
C 5	23	57.5	161652	11	AC004774	AC004774 Homo sapi
C 6	23	57.5	103911	11	HDAC002565	AC002565 Human Chr
C 7	22.8	57.0	18227	12	AF110420	AF110420 Mus muscu
C 8	22.4	56.0	105160	8	AC002328	AC002328 genomic s
C 9	22.4	56.0	110858	10	HUM263A	AC000062 Human PAC
C 10	22.4	56.0	170215	34	AC004127	AC004127 Homo sapi
C 11	22.4	56.0	200965	34	AC006740	AC006740 Caenorhab
C 12	22.4	56.0	37678	35	CEIK0812	U97001 Caenorhabd
C 13	22.2	55.5	236496	34	AC006790	AC006790 Caenorhab
C 14	22.2	55.5	19566	37	CE149F8C	AF123453 Caenorhab
C 15	22	55.0	3498	5	A37232	A37232 Sequence 1
C 16	22	55.0	100000	9	AP000084	AP000084 Homo sapi
C 17	22	55.0	100000	9	AP000136	AP000136 Homo sapi
C 18	22	55.0	89323	9	AP000221	AP000221 Homo sapi
C 19	22	55.0	274498	34	AC006765	AC006765 Caenorhab
C 20	22	55.0	110000	34	CEY11B2_4	Continuation (5 of
C 21	22	55.0	110000	34	CEY37D8_1	Continuation (2 of
C 22	22	55.0	291692	34	CEY59A8	298870 Caenorhabd
C 23	22	55.0	85316	35	AC005978	AC005978 Drosophi
C 24	22	55.0	41299	35	CEC06H5	292775 Caenorhabd
C 25	22	55.0	132742	35	CEY37D8A	AL032626 Caenorhab
C 26	22	55.0	16938	37	CE1725D3	AF077539 Caenorhab
C 27	21.8	54.5	175818	11	AC006257	AC006257 Homo sapi
C 28	21.6	54.0	9639	2	U67483	U67483 Methanococ
C 29	21.6	54.0	99053	8	AC005292	AC005292 genomic s
C 30	21.6	54.0	177726	11	AC004083	AC004083 Homo sapi
C 31	21.6	54.0	330250	11	AF048895	AF048895 Homo sapi
C 32	21.6	54.0	43888	34	CEY43E12	298862 Caenorhabd
C 33	21.6	54.0	29731	36	CEB0391	281454 Caenorhabd
C 34	21.6	54.0	32320	35	CEY13H10	268748 Caenorhabd
C 35	21.6	54.0	2732	35	HVU24116	U24116 Hydra vulga
C 36	21.6	54.0	106908	35	PEMAL3P8	AL034560 Plasmodiu
C 37	21.4	53.5	3484	8	AE063234	AE063234 pneumocys
C 38	21.4	53.5	152592	9	HS396D17	AL008634 Human DNA
C 39	21.4	53.5	121424	9	HS418A9	Z84480 Homo sapien
C 40	21.4	53.5	117435	11	HUAC003964	AC003964 Homo sapi
C 41	21.4	53.5	226683	35	AC007297	AC007297 Homo sapi
C 42	21.4	53.5	40114	35	CE12K6	AF056942 Caenorhab
C 43	21.2	53.0	30389	7	SPAC19D5	Z99531 S.pombe chr
C 44	21.2	53.0	44092	7	SPAC6F6	Z99581 S.pombe chr
C 45	21.2	53.0	110000	34	CEY81G3_0	AL021152 Caenorhab

ALIGNMENTS

RESULT 1
HSJ661G12/c
LOCUS Homo sapiens chromosome 1 clone J661G12, WORKING DRAFT SEQUENCE. In
DEFINITION Homo sapiens chromosome 1 clone J661G12, WORKING DRAFT SEQUENCE. In
ACCESSION AL049797
NID G5051384
VERSION AL049797.2 GI:5051384
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 185092)
AUTHORS Frankland, J.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@anger.ac.uk Clone requests: clonerequests@anger.ac.uk On Jun 11, 1999 this sequence version replaced gi:4775448.

COMMENT IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be

contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known: 800 n's separate segments. Unfinished: d661G12 Contig_ID: 02987 acc=AL049797 Length: 123411 bp Unfinished: d661G12 Contig_ID: 02328 acc=AL049797 Length: 15975 bp Unfinished: d661G12 Contig_ID: 00952 acc=AL049797 Length: 44106 bp.
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES
 source
 location/Qualifiers
 1. 185092
 /db_xref="taxon:9606"
 /chromosome="1"

BASE COUNT 56011 a 32549 c 34225 g 60707 t 1600 others
 ORIGIN

Query Match 65.0%; Score 26; DB 34; Length 185092;
 Best Local Similarity 85.3%; Pred. No. 10;
 Matches 29; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 AATTCATGCTGTTTCATGCAACAATCA 40
 |||
 Db 169799 AATTCATGCTGTTTCATGCAACAATCA 169766

RESULT 2
 AC005354/c
 LOCUS Homo sapiens chromosome 5, p1 clone 1320c6 (UBNL H59), complete
 DEFINITION
 AC005354 AC002244 AC001059 AC001060 AC001061 AC001536 L81878 L81879
 L81880 L81881
 93366558
 AC005354.1 GI:3366558
 HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
 AUTHORS Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
 1 (bases 1 to 28065)
 Kaderer,K., Miguel,T., Miller,C., Piltuck,S., Pollard,M.,
 Rojeski,H., Subramanian,S. and Martin,C.H.
 Sequencing of human chromosome 5
 Unpublished

JOURNAL
 REFERENCE
 AUTHORS 2 (bases 1 to 28065)
 Rieke,D.O.
 Title Large Scale Sequence Analysis and Annotation with the Sequence
 Comparison Analysis (SCAN) System
 Unpublished

JOURNAL
 REFERENCE
 AUTHORS 3 (bases 1 to 28065)
 Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
 Davis,C.A., Kaderer,K., Miguel,T., Piltuck,S., Pollard,M.,
 Rojeski,H., Subramanian,S. and Martin,C.H.
 Direct Submissio

TITLE
 JOURNAL Submitted (31-JUL-1998) Human Genome Center, DOE Joint Genome
 Institute, Lawrence Berkeley National Laboratory, MS 74-157,
 Berkeley, CA 94720, U.S.A.
 Sequence submitted by:
 DOE Joint Genome Institute.

COMMENT
 FEATURES
 source
 location/Qualifiers
 1. 28065
 /db_xref="taxon:9606"
 /map="5q"
 /clone="1320c6"
 /chromosome="5"
 /note="UNL H59"

repeat_region 98..481

repeat_region 491..916
 /rpt_family="L1"
 /rpt_family="Alu"
 1158..1285
 /rpt_family="MER42"
 complement(1301..1700)
 /standard_name="LTR7"
 complement(1598..2328)
 /rpt_family="MER39"
 complement(2002..2135)
 /rpt_family="MER21"
 complement(2340..2625)
 /rpt_family="Alu"
 2882..3031
 /rpt_family="MER42"
 3436..3546
 /rpt_family="Alu"
 3537..3560
 /note="(A)24"
 /rpt_type=tandem
 /rpt_unit=A
 3955..4060
 /rpt_family="Alu"
 complement(4675..4767)
 /rpt_family="MIR"
 4308..5061
 /rpt_family="MIR"
 complement(5233..5288)
 /rpt_family="Alu"
 5760..6044
 /rpt_family="Alu"
 complement(join(6081..6266,15732..15929,22197..22366))
 /standard_name="novel protein kinase"
 /note="53-83% identity p56 protein kinase U35146"
 6545..6669
 /rpt_family="Alu"
 complement(7302..7598)
 /rpt_family="Alu"
 complement(8141..8567)
 /rpt_family="Alu"
 complement(8586..8837)
 /rpt_family="L1"
 complement(9393..9686)
 /rpt_family="Alu"
 complement(10400..10474)
 /rpt_family="LINE2"
 10505..11088
 /rpt_family="Alu"
 11074..11094
 /note="(A)21"
 /rpt_type=tandem
 /rpt_unit=A
 complement(11759..12270)
 /rpt_family="MER1"
 12341..12626
 /rpt_family="Alu"
 13002..13478
 /rpt_family="Alu"
 13778..13938
 /rpt_family="Alu"
 14432..14455
 /note="(ATT)6"
 /rpt_type=tandem
 /rpt_unit=ATT
 complement(14611..14694)
 /rpt_family="Alu"
 14864..15141
 /rpt_family="Alu"
 complement(15735..15929)
 /note="GRAIL 2 excellent exon, frame 1"
 complement(16047..16302)
 /rpt_family="Alu"
 complement(17088..17212)


```

repeat_region      6113. .6662
                    /rpt_family="MER41A"
repeat_region      complement(6663. .6994)
                    /rpt_family="MER83"
repeat_region      complement(7000. .7557)
                    /rpt_family="MER4. internal"
repeat_region      complement(7560. .8234)
                    /rpt_family="MER4A"
repeat_region      8239. .9165
                    /rpt_family="MSTB-internal"
repeat_region      complement(9215. .9483)
                    /rpt_family="AluJo"
repeat_region      9484. .9748
                    /rpt_family="MSTB-internal"
repeat_region      9841. .10206
                    /rpt_family="MSTB"
repeat_region      complement(10291. .10484)
                    /rpt_family="AluJo"
repeat_region      complement(10495. .10794)
                    /rpt_family="AluSg"
repeat_region      11777. .12076
                    /rpt_family="AluSx"
repeat_region      12077. .12105
                    /rpt_family="AluSx"
repeat_region      12479. .12775
                    /rpt_family="(TAA)n"
repeat_region      complement(13041. .13531)
                    /rpt_family="LINE2"
repeat_region      13809. .13944
                    /rpt_family="AluSx"
repeat_region      13945. .14247
                    /rpt_family="AluSg"
repeat_region      14248. .14417
                    /rpt_family="AluSx"
repeat_region      14504. .14618
                    /rpt_family="AluSg1"
repeat_region      14620. .14895
                    /rpt_family="AluSg"
repeat_region      14896. .15064
                    /rpt_family="AluSg1"
repeat_region      15118. .15404
                    /rpt_family="AluSg"
misc_feature       15334. .15566
                    /note="DSD similarity to overlapping ESTs:-(15334. .15684)
                    AA013249. ze26c07.s1 Soares retina N2b4HR Homo sapiens cDNA
                    clone 360108 3' similar to gp:M35663 INTERFERON-INDUCED,
                    DOUBLE-STRANDED RNA-ACTIVATED PROTEIN KINASE
                    (HUMAN); contains Alu repetitive element; Score: 645
                    identity: 359/370 (97%)--(15334. .15566) AA016313
                    ze38h03.s1 Soares retina N2b4HR Homo sapiens cDNA clone
                    361301 3' similar to gp:M35663 INTERFERON-INDUCED,
                    DOUBLE-STRANDED RNA-ACTIVATED PROTEIN KINASE
                    (HUMAN); contains Alu repetitive element; Score: 449
                    identity: 230/232 (99%)."
repeat_region      16868. .16930
                    /rpt_family="MLT1A2"
repeat_region      16932. .16978
                    /rpt_family="POLY-A"
repeat_region      complement(17033. .17060)
                    /rpt_family="AT_rich"
misc_feature       17101. .17239
                    /note="DSD similarity to AA054173 z5f1a08.r1 Soares retina
                    N2b4HR Homo sapiens cDNA clone 380438 5' similar to
                    contains Alu repetitive element; (416. .277); 99%
                    identity."
repeat_region      17188. .17319
                    /rpt_family="AluJo"
repeat_region      17340. .17797
                    /rpt_family="AluSg"
misc_feature       17714. .18004
                    /note="DSD similarity to overlapping ESTs:-(17728. .18004)
                    AA054173 z5f1a08.r1 Soares retina N2b4HR Homo sapiens cDNA
                    clone 380438 5' similar to contains Alu repetitive

```

```

element; (276. .1); 99% identity.-(18004. .17714) A013121
ze26c07.r1 Soares retina N2b4HR Homo sapiens cDNA clone
360108 5' similar to contains Alu repetitive element;
(1. .292); 97% identity.-(18004. .17714) AA017502
ze38h03.r1 Soares retina N2b4HR Homo sapiens cDNA clone
361301 5' similar to contains Alu repetitive element;
Score: 567 identity: 292/297 (98%)."
repeat_region      17892. .18191
                    /rpt_family="AluJb"
repeat_region      complement(18516. .18710)
                    /rpt_family="AluJo"
repeat_region      18906. .19203
                    /rpt_family="AluY"
repeat_region      19288. .20021
                    /rpt_family="LINE2"
repeat_region      20258. .20313
                    /rpt_family="LINE2"
repeat_region      complement(20739. .21323)
                    /rpt_family="LTR10C"
repeat_region      complement(21324. .21361)
                    /rpt_family="MIR"
repeat_region      21365. .21487
                    /rpt_family="FLAM_A"
repeat_region      21491. .21793
                    /rpt_family="AluJb"
repeat_region      complement(21817. .21953)
                    /rpt_family="MIR"
misc_feature       22765. .22991
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 0, quality: excellent, score: 86.000"
repeat_region      22810. .22867
                    /rpt_family="(GAA)n"
misc_feature       24453. .24713
                    /note="DSD similarity to A362319 EST95595 Testis I Homo
                    sapiens cDNA 5' end. Score: 505 identity: 258/260 (99%)."
misc_feature       25318. .25887
                    /note="DSD similarity to AA061824 m133d04.r1 Stratagene
                    mouse testis (#937308) Mus musculus cDNA clone 513799 5';
                    Score: 279 identity: 295/437 (67%)--(25480. .25887)
                    BLASTN similarity to AA860210 (1. .407); match: 1, score:
                    5.7e-157; database searched: month.na; ak48c12.s1 Soares
                    testis NHT Homo sapiens cDNA clone IMAGE:1409206 3'"
repeat_region      complement(26437. .26479)
                    /rpt_family="MLT1D"
repeat_region      26514. .26804
                    /rpt_family="AluJb"
repeat_region      complement(26831. .27035)
                    /rpt_family="MLT1D"
repeat_region      complement(27042. .27333)
                    /rpt_family="AluY"
repeat_region      complement(27334. .27637)
                    /rpt_family="AluSp"
repeat_region      complement(27644. .27716)
                    /rpt_family="MLT1D"
repeat_region      27727. .28020
                    /rpt_family="AluSx"
repeat_region      28681. .28972
                    /rpt_family="AluSx"
repeat_region      29177. .29471
                    /rpt_family="7SLRNA"
misc_feature       complement(29214. .29339)
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 0, quality: good, score: 50.000"
Query Match      58.0%; Score 23.2; DB 11; Length 86765;
Best Local Similarity 77.8%; Pred. No. 87;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY      1 TGCTGAATTCCTAATGCTGATTTTCATGAACAA 36
Db      11706 TGCTGACATTCAAATGCTGACATGACAGACTA 11671

```



```

repeat_region 7510..7596
/rpt_family="MER1_type"
repeat_region 7809..7910
/rpt_family="L2"
repeat_region 8148..8256
/rpt_family="L2"
repeat_region 8480..8667
/rpt_family="L1"
repeat_region 9112..9205
/rpt_family="MER1_type"
repeat_region 11511..11804
/rpt_family="Alu"
repeat_region 13624..13993
/rpt_family="MALR"
repeat_region 14122..14252
/rpt_family="MIR"
repeat_region 20744..21164
/rpt_family="L1"
repeat_region 21166..21314
/rpt_family="L1"
repeat_region 21331..21614
/rpt_family="Alu"
repeat_region 21615..21664
/rpt_family="(CATR)n"
repeat_region 21665..22121
/rpt_family="L1"
repeat_region 22171..22237
/rpt_family="MIR"
repeat_region 23047..23347
/rpt_family="Alu"
repeat_region 23357..23718
/rpt_family="MALR"
repeat_region 24659..24960
/rpt_family="Alu"
repeat_region 25327..25559
/rpt_family="L1"
repeat_region 25826..25949
/rpt_family="L1"
repeat_region 29100..29380
/rpt_family="Alu"
repeat_region 29385..29482
/rpt_family="(TAGA)n"
repeat_region 29544..29646
/rpt_family="L2"
repeat_region 30131..30242
/rpt_family="L2"
repeat_region 30355..30653
/rpt_family="Alu"
repeat_region 30948..31107
/rpt_family="MER1_type"
repeat_region 33110..34309
/rpt_family="MER2_type"
repeat_region 34416..34739
/rpt_family="L2"
repeat_region 35369..35571
/rpt_family="L2"
repeat_region 35364..35798
/rpt_family="L2"
repeat_region 35818..36225
/rpt_family="L2"
repeat_region 36354..37334
/rpt_family="L1"
repeat_region 37282..37409
/rpt_family="MIR"
repeat_region 38220..38518
/rpt_family="Alu"
repeat_region 39945..40366
/rpt_family="L2"
repeat_region 41772..42169
/rpt_family="MALR"
repeat_region 42170..42339
/rpt_family="MIR"
repeat_region 42978..43147

```

```

repeat_region 43148..43639
/rpt_family="MIR"
repeat_region 44708..45006
/rpt_family="Retroviral"
repeat_region 45940..46256
/rpt_family="Alu"
repeat_region 46330..46564
/rpt_family="7SLRNA"
repeat_region 48279..48540
/rpt_family="L2"
repeat_region 48927..49030
/rpt_family="Alu"
repeat_region 49237..49527
/rpt_family="L1"
repeat_region 49528..49613
/rpt_family="Alu"
repeat_region 50590..50877
/rpt_family="(TAGA)n"
repeat_region 50890..51025
/rpt_family="Alu"
repeat_region 53705..53841
/rpt_family="MIR"
repeat_region 54158..54335
/rpt_family="MIR"
repeat_region 54676..54750
/rpt_family="MIR"
repeat_region 57756..57891
/rpt_family="L2"
repeat_region 62590..62636
/rpt_family="MIR"
repeat_region 63046..63203
/rpt_family="L2"
repeat_region 63324..64041
/rpt_family="L1"
repeat_region 64113..64173
/rpt_family="L2"
repeat_region 64181..64443
/rpt_family="L2"
repeat_region 64506..65064
/rpt_family="Alu"
repeat_region 66287..66581
/rpt_family="L2"
repeat_region 67183..67278
/rpt_family="Alu"
repeat_region 67515..67605
/rpt_family="MIR"
repeat_region 72685..72966
/rpt_family="MIR"
repeat_region 72685..72966
/rpt_family="Alu"

```

```

Query Match 57.5% Score 23; DB 11; Length 161652;
Best Local Similarity 74.4% Pred. No. 91;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 TGCTGTAATCTATGCTGCTATTTCATGAACAATG 39
Db 91650 TGCTTAATGTAATGTAATGTAATGTAATGTAATG 91612

```

```

RESULT 6
HDCUS HDAC002565 103911 bp DNA PRI 17-MAR-1998
LOCUS HDAC002565 16 BAC clone CIT987SK-A-598D4, complete sequence.
DEFINITION Human Chromosome 16
ACCESSION AC002565
NID 92896804
VERSION AC002565.1 GI:2896804
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 103911)
Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
AUTHORS

```


CDS
join(825..1085,1254..1745,1901..2397,2519..2743,
2884..3112)
/evidence-not_experimental
/codon_start-1
/product-"F20N2.1"
/protein_id-"AAC83053.1"
/db_xref="GI:3953456"
/db_xref="GI:3953456"
/translation-"MEESRDILMTLIESGVSIPIGDTVSSETPPALVSICAOILLN
LIDPSASEDELDPSLEPERRICDIASHVKNLKYINDMSYKKNVLFKTLNGLGCGS
YLCEHAEETOCMCILROFLHSEEDSRVLFVLEYSIESEKRTLAGIAPKPK
ETPRDISDMDVNRDKDETPDMHIOKVEAVLKLTLMTSEKSSDSLAKNTSAVDS
SOKTDDPVTNRDLSLSDSCREENSEDEPFTNVEYELONOHDLLEESGSSO
LCSLESELLOAAERLDDKRGGSYLDLNOQLVYKRONIDILKORTQYITIF
KLASHSVSNITSGHDGRVMTITVIMSSFRDQVLETKLLLDQLHAVEEP
AKERHKLRTLELDQSLSEIORERECRLYNELRPPAPRKSYPHIOIKETRN
SRKIDITQISGETRELQLEKNSIOERLSRYAVVDMETREVKRDPAAQVYKLLT
SIHIFQISEKILMTDRFRETIVDYENKLGSLTARQMSLEKIQADDAIKENESLK
K"
CDS
complement(join(3612..3905,3995..4372,4466..4715,
4916..5204,5303..5406,5505..5625,5711..5786,5878..5927,
6459..6633))
/note-"hypothetical protein"
/codon_start-1
/evidence-not_experimental
/product-"F20N2.3"
/protein_id-"AAC83053.1"
/db_xref="PID:93953457"
/db_xref="GI:3953457"
/translation-"MSSSDGSSSVSIPIYHLKTLQIKREYTGKSHDEDFEAYKDS
FNDPHEAOKLFLDTPHEVRSREKKEKASPIYVQPSGGRGRNPASNSIQGSS
GRNASFRENGANHVTGSRTPAQPAIKASITVFNKTSVGPASIPSEVSNHQAOD
PSLISARCSKSDQALELTASQKGNQSLPKPDVSESHVTPFHLQVAKGLQNGL
TFSGDFNFPKEVSSNGASGGYSDNESHGTDDESPITNGITVYASAEETS
TVSEDKDYGISNSATGAEPYVSDHIVPEVEYKEBALSNTEHQLAYGEALSYF
GLVPSLAIQGPVTEAEPQPGNSGSPISLIVYPOSSIAATOOTNLTRQOYVP
NFPYGVSYVMPPIYHOELSPNGIPQOSYFGAALTRAPSHAKPYDNTENPTTN
PYLHSPVNASITSTTTLNSIHSEKASHTESAAVMTGGFENLQVNPMTAYOG
QPLGFVYQAGHGMHOPOTPMAASTYOTLPPPHTTMGEPIGHPIAYOO
POAALTMVNNY"
CDS
complement(join(7027..7176,7296..7378,7472..7704,
7774..7834,7942..8002,8076..8159,8252..8446,8523..8602,
8649..8832,9045..9163,9276..9345,9447..9538,9604..9757))
/note-"unknown: similar to uridine kinase (Z99117);
similar to uracil phosphoribosyl transferase (A1022223);
similar to ESTs gb|746383, gb|N99614, and gb|AA979998"
/codon_start-1
/evidence-not_experimental
/product-"F20N2.3"
/protein_id-"AAC83054.1"
/db_xref="PID:93953458"
/db_xref="GI:3953458"
/translation-"MASKSDVNIETSSKVFSGFHQMDGLASNRPRQMAEEHGGP
FVIESRGVAGASGKTVCMDIMQOLHQRVAVVNDSPYHNVNEIYRVADYDP
DHDPADPEQLISMERLKGQOAVDIPNIDSKYKNINPEPRRNPSSVITILELIF
HDPVRLAMKRIFYDAGLSHTKPVNTVYVAVAKRCTICTHEDEHVAIDLVYHIRT
DYEKGRDIATLDQYSKVPKPAEDFILPKKVAIDILIPGSGNHAVIDLVYHIRT
KLQGDJCKTIPNLVYOSTFOIINGMTLLIDSSTKTHDFEYSDRLIRLVYEGHGH
LPTEHQVVTPTGSVYGDFCKLGCYSVSRSESEHNAIPACGKIKIKILHRE
GDNGOQCVSLITSPNYLTNGTHQLYERLPSDISEHVALLDPIIGTSGSAVO
AIRLLISGVPESSIIIFNLISAPGVAVYCKPRIKIYVSELEGLNDEFRRVPCG
GEGDRIFGIDDEP"
11592..12074
/note-"unknown: similar to A1022223; similar to EST
gb|N37483"
/codon_start-1
/evidence-not_experimental
/product-"F20N2.4"
/protein_id-"AAC83055.1"
/db_xref="PID:93953459"
/db_xref="GI:3953459"
/translation-"MSSSIRLIVSGFHRQPLKSPVNSPVSFISVPEFMSKSG
TGRSVAMSSVEKTSQSGALENRASRREKLEQLIPVELIVEDYVSIQAHGAGMKC

CDS
join(12795..12975,13124..13403,13459..14176)
/note-"hypothetical protein: similar to A1022223"
/codon_start-1
/evidence-not_experimental
/product-"F20N2.5"
/protein_id-"AAC83056.1"
/db_xref="PID:93953460"
/db_xref="GI:3953460"
/translation-"MRELSELNRDQDEEYVWVYKSHQILLVGEDESECSLAT
CFGASNIVASLSDSYKPVNCCSFRLDPLSCGSMFVITKPDVYRKKNASNL
TLKLGAFLLHGADATLLHHPDRLRYRPRVITNPFHTGPHKRESPCOIQPAATL
TNLRFDFLHGASHMLRADGEVYHSHKKAPECYWNEELASRCFLVILQLEAKRN
PGYENKKGDSRLCQDPLLGECSTYKFRFRAKELVAKRYEVEKCPRLSTRAN
KHAREFSISHLDPCTCFKFKVSIEMHNGELQVRYITRFPNHYTKSSETRRL
CODVAVQASQEPFERBLRRFGVSHGHNGRVRKMLIRNSTGEERKLRMRQRLKIR
"
CDS
join(14961..15147,15440..15638,15767..16253,16466..16690,
16860..16958,17122..17236)
/note-"similar to A1022223 and AF067215"
/codon_start-1
/evidence-not_experimental
/product-"F20N2.6"
/protein_id-"AAC83057.1"
/db_xref="PID:93953461"
/db_xref="GI:3953461"
/translation-"MAMRELSQLNRSDQDEEYVWVYKSHQILLVGEDESECSL
ATLGASNIVASLSDSYKPVNCCSFRLDPLSCGSMFVITKPDVYRKKNASNL
YRRDRYFNPNGAFGRGSDSLTKRHHLYNGFNGASRLLRANGEVYHKNNA
PFSMNLELASRCFLVILQVAPAEKNITPEYENKRDGRCDPFLIGESTYKFR
SVRAKELYAEVRERERESMPEALINKQVSPDRRLQTEFGVSGQRPVLFPL
SYRDNLRQVQDPLFHESSYRTRVSPDIYDELIQMLTRTSPPHYAGSPERHRL
LCQFPOVASQEPSEYSNRMSGVSLIYMEVYILKNSGYDGKVSBAWDSISI
SKNVEITIEFVALLLEDHRSVSDSLDITWTGDLNLSVCF"
join(17736..17910,18005..18229,19477..19991)
/note-"hypothetical protein"
/codon_start-1
/evidence-not_experimental
/product-"F20N2.7"
/protein_id-"AAC83058.1"
/db_xref="PID:93953462"
/db_xref="GI:3953462"
/translation-"WOLFSPHLEFKNLVSKLVSCKTRNITFSFLQORIVLKMSDE
KSNKRAKIASKPGVSVSIQGNODLVLEGEIDLAETRELKRYCMTTITTYVA
APQOPQPHMGQYQMPPARCTCIPNAGVLFICIVLISVYTNANSGMSDLC
KHSDPMLCISITSRPESGEFACTSNOETIATSAANASATSYITKOKLSMED
PAIEDLQDOKYQDAVBOLDISLMMADHIVDVWLSAATSAIESGSLGSA
GNDLELSQNRNVLKLCNMLINKMLT"
join(22523..22641,22783..22931,23048..23166,23311..23913)
/note-"similar to MEL-26 (U67737)"
/codon_start-1
/evidence-not_experimental
/product-"F20N2.8"
/protein_id-"AAC83059.1"
/db_xref="PID:93953463"
/db_xref="GI:3953463"
/translation-"MTDSAYRVDTISRLAQRIHNLSSSTRKSDPFRMGLMNLHLSV
EKSMLNVLKYPVSNLITRENPPVAFALRSTSEKRALSHPEYIDRKIKNEP
IMTEVLELIGITIIDVFDLKYLSQSGSELYSWANGSTENQOYAVYSGMLTE
SIYDITINSDSIGAHRAVLAKRSVFSMFLHDLKEKELSTINVLQNLQCAF
LSYVGINQMEDFLHRLALQAAEKYDIALDKACHSLDLDITDYNVERLQNAVL
YQPLKASQCRVYKFKGKIFEIDENINMOCDRLISIEFEVYSTWKG"
complement(join(24535..24616,24851..24962,25135..25159))
/note-"hypothetical protein"
/codon_start-1
/evidence-not_experimental
/product-"F20N2.9"
/protein_id-"AAC83060.1"
/db_xref="PID:93953464"
/db_xref="GI:3953464"
/translation-"MSTAAPEEASVEYQYLSISKLEEMKRISSKKKKAMVYVIGSC
RSPFALKRISIEVLESDFVSTIEDLKSIL"


```
repeat_region 56261..56547
/rpt_family="ALU"
misc_feature 57853..58519
/gene="WUGSC:H_2G3A.1"
/note="match to human 5' EST T77253 (NID:9694456) and 3'
EST T77470 (NID:9694673), reads from opposite ends of the
same clone"
exon complement(58576..58620)
/gene="WUGSC:H_2G3A.1"
/note="Grail prediction, score = 96"
/evidence=not_experimental
repeat_region 59652..59931
/rpt_family="ALU"
63783..64081
/rpt_family="ALU"
69913..70050
/gene="WUGSC:H_2G3A.1"
/note="Grail prediction, score = 88"
/evidence=not_experimental
repeat_region complement(73248..73536)
/rpt_family="ALU"
74739..75028
/rpt_family="ALU"
repeat_region complement(76412..76709)
/rpt_family="ALU"
77367..77447
exon /gene="WUGSC:H_2G3A.1"
/note="Grail prediction, score = 94"
/evidence=not_experimental
78417..78533
/gene="WUGSC:H_2G3A.1"
/note="Grail prediction, score = 80"
/evidence=not_experimental
repeat_region 80069..80361
/rpt_family="ALU"
complement(82638..82912)
/rpt_family="ALU"
repeat_region complement(83396..83688)
/rpt_family="ALU"
84245..84412
exon /gene="WUGSC:H_2G3A.1"
/note="Grail prediction, score = 96"
/evidence=not_experimental
repeat_region complement(86357..86650)
/rpt_family="ALU"
repeat_region complement(98118..98415)
/rpt_family="ALU"
repeat_region complement(98916..99192)
/rpt_family="ALU"
99855..99956
exon /gene="WUGSC:H_2G3A.1"
/note="Grail prediction, score = 82"
/evidence=not_experimental
100410..100538
exon /gene="WUGSC:H_2G3A.1"
/note="Grail prediction, score = 98"
/evidence=not_experimental
105514..105666
exon /gene="WUGSC:H_2G3A.1"
/note="Grail prediction, score = 83"
/evidence=not_experimental
108897..109182
repeat_region /rpt_family="ALU"
BASE COUNT 33497 a 20778 c 21335 g 35248 t
ORIGIN
Query Match 56.0%; Score 22.4; DB 10; Length 110858;
Best Local Similarity 72.5%; Pred. No. 1.5e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
DB 27259 TCCAAAAATTTTATGCGGCTTTTATACACAAATTA 27298
RESULT 10
AC004127 170215 bp DNA HTG 06-FEB-1998
LOCUS Homo sapiens chromosome 11 clone pD741n15, WORKING DRAFT SEQUENCE.
DEFINITION 7 unordered pieces.
ACCESSION AC004127
NID 92842777
VERSION AC004127.1 GI:2842777
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170215)
Evans, G.A., Athanasiou, M., Basil, M., Bradbury, P., Brignac, S.,
Bunester, R., Davis, C., English, C., Franklin, T.L., Garner, H.R.,
Gee, V., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J.,
Hinson, S., Narayanaswamy, U., Newton, J., O'Brien, K., Patel, P.,
Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D.,
Ward, T. and Wilson, R.
TITLE template
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170215)
Evans, G.A., Athanasiou, M., Basil, M., Bradbury, P., Brignac, S.,
Bunester, R., Davis, C., English, C., Franklin, T.L., Garner, H.R.,
Gee, V., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J.,
Hinson, S., Narayanaswamy, U., Newton, J., O'Brien, K., Patel, P.,
Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D.,
Ward, T. and Wilson, R.
DIRECT SUBMISSION
TITLE University of Texas Southwestern Medical Center, 5323 Harry Hines
Blvd, Dallas, TX 75235-8591, USA
JOURNAL Submitted (06-FEB-1998) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines
Blvd, Dallas, TX 75235-8591, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2519: contig of 2519 bp in length
* 2 4982: contig of 2463 bp in length
* 3 4983 13767: contig of 8785 bp in length
* 4 13768 32352: contig of 18585 bp in length
* 5 32353 55696: contig of 23344 bp in length
* 6 55697 86119: contig of 30423 bp in length
* 7 86120 170215: contig of 84096 bp in length.
FEATURES
Location/Qualifiers
source 1..170215
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="pD741n15"
/chromosome="11"
BASE COUNT 47532 a 37634 c 37608 g 47376 t 65 others
ORIGIN
Query Match 56.0%; Score 22.4; DB 34; Length 170215;
Best Local Similarity 72.5%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```

DEFINITION  Caenorhabditis elegans clone Y38B5, WORKING DRAFT SEQUENCE, 12
ACCESSION   AC006740
NID         94309800
VERSION     AC006740.2  GI:4309800
KEYWORDS    HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
ORGANISM    Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
            Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 200965)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Caenorhabditis elegans clone
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 200965)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (23-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 1, 1999 this sequence version replaced g1:4263198.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 33116: contig of 33116 bp in length
            * 33117 33125: gap of unknown length
            * 33126 36942: contig of 3817 bp in length
            * 36943 36951: gap of unknown length
            * 36952 39003: contig of 2052 bp in length
            * 39004 39012: gap of unknown length
            * 39013 46010: contig of 6998 bp in length
            * 46011 46019: gap of unknown length
            * 46020 59290: contig of 13271 bp in length
            * 59291 59299: gap of unknown length
            * 59300 130771: contig of 71472 bp in length
            * 130772 130780: gap of unknown length
            * 130781 155202: contig of 24422 bp in length
            * 155203 155211: gap of unknown length
            * 155212 160656: contig of 5435 bp in length
            * 160657 160675: gap of unknown length
            * 160676 170159: contig of 9484 bp in length
            * 170160 170168: gap of unknown length
            * 170169 173251: contig of 3083 bp in length
            * 173252 173260: gap of unknown length
            * 173261 176650: contig of 3390 bp in length
            * 176651 176658: gap of unknown length
            * 176659 200965: contig of 24307 bp in length.
            *
            * Location/Qualifiers
            *   1..200965
              /organism="Caenorhabditis elegans"
              /db_xref="taxon:6239"
              /clone="Y38B5"
BASE COUNT  61403 a 37295 c 36272 g 64715 t 1280 others
ORIGIN
Query Match      56.0%; Score 22.4; DB 34; Length 200965;
Best Local Similarity 72.5%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

DY 1 TGCCTAATTCATGCTGATTCATGATCAATGA 40
    ||||| ||||| ||||| ||||| |||||
Db 109845 TTCTGAAATTTTATCTGCAATTCATGACGCGTGA 109806

RESULT  12
CELK08B12/C
LOCUS   CELK08B12 37678 bp DNA INV 16-APR-1997

```

```

DEFINITION  Caenorhabditis elegans cosmid K08B12.
ACCESSION   U97001
NID         91938418
VERSION     U97001.1  GI:1938418
KEYWORDS
SOURCE      Caenorhabditis elegans strain-Bristol N2.
ORGANISM    Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
            Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 37678)
AUTHORS     Wilson,R., Altschough,R., Anderson,K., Baynes,C., Berts,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,E., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Lalster,N.,
            Latreille,P., Lightning,J., Lloyd,C., McKurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Roopra,A.,
            Saunders,D., Showkeen,R., Smaildon,N., Smith,A., Sonhammer,E.,
            Staden,R., Suston,J., Thierly-Mieg,U., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Welstock,L.,
            Wilkinson-Spratt,J. and Wohlman,P.
            2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans.
            Nature 368 (6466), 32-38 (1994)
            94150718
REFERENCE   2 (bases 1 to 37678)
AUTHORS     Becker,M. and Wohlmann,P.
TITLE       The sequence of C. elegans cosmid K08B12
JOURNAL     Unpublished (1997)
REFERENCE   3 (bases 1 to 37678)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (10-APR-1997)
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63110, USA, and
            Sanger Centre, Hinxton Hall
            Cambridge CB10 1RO, England
            e-mail: iwenematode.wustl.edu and jesssanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' end of this cosmid lies in gap:3' cosmid is F31F7, 7400 bp
overlap. Actual start of this cosmid is at base position 1 of
CELK08B12; actual end is at 33279 of CELK08B12

NOTES:
Coding sequences below are predicted from computer analysis, using
the program GeneFINDER (P. Green and L. Hillier, ms in preparation).
Location/Qualifiers
1..37678
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="V"
/clone="K08B12"
5609..8033
gene

```

gene
CDS
/gene="K08B12.3"
join(5609..5746,5903..6062,7298..7396,7446..7558,
7614..7791,7948..8033)
/gene="K08B12.3"
/note="similar to Schistosomaccharomyces pombe
4-nitrophenylphosphatase (PNPPASE) (GB:62722, NID:g5005)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB52259.1"
/db_xref="GI:1938421"
/db_xref="PIR:1938421"
/translation="MSKISSVLIDSGTTHIEEPATPGQATLELRQAKVFTVNT
TKSKRLHORLINCGRKVEKEFTSLTARLIVNQYRPFYIYDNRAMEDEGIS
TDPNNAVIGLAPRENDITLHAFRLIKRKLINRGTHQINAGLCIGGITY
AGLEYSAGVATVGRKPLFEFALOSLLENVDESSAVMIGDVEDDALGAIKIGAR
AIIYKGRFDRGDELKRYKVAANSFVDAVNMIIENKENG"
18322..22155
/gene="K08B12.2"
join(18322..18430,19598..19742,20024..20165,20733..20994,
21048..21779,21947..22155)
/gene="K08B12.2"
/note="short region of similarity to the DNA-binding
domain of drosophila doublesex protein (SP:P23022,
NID:g475979)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB52258.1"
/db_xref="PIR:g1938420"
/db_xref="GI:1938420"
/translation="MYPRSYGSLSHMTENGHOPEVYHSHSLNKERYDLYRALIGPL
VVASLOYHPRGFESLVCRCICFVWLANEVRVTEIMEIYATIPQSYVLEAO
LITSVDHTEIEVQEVQEQRLNGGLTCGASGSSRRITLCRCSEHGQVVLKGA
SRPFPNCSCKTCTVWMSRANAIIIRRYRTLEGGLVLPVHEKNGTILRYEPKNI
DESDALTIHFNKNONGLANMANNAEMQMGAVPQSPATAPDSGNAQMPKYS
OSDOELRGHPTIYECAVOOQESATNPYGLDQSPMPPESTANGOMPMGL
SDPSLTPETITAKNLNPGNDIVTTTPSIPFTELSNLTANIGVYTSANN
GNOMLTPQYTTASTGMPFEDIDRLSNGDSMLIPANSLNDLNIANSNTNKT
TELAPKLPHEMTDACLITNGDSPEIKETMEPIQLNLLSLIDPSRHPNF
QMFIDCVSLKTMLEPSSSTENST"
complement(28456..37616)
/gene="K08B12.5"
complement(join(28456..28488,29203..29350,29407..29604,
31204..31495,32179..32335,32386..33139,33265..34447,
34683..35391,35462..35801,36273..36537,36707..36978,
37370..37450,37502..37616))
/gene="K08B12.5"
/note="Similar to serine/threonine-protein kinase; coded
for by C. elegans cDNA CEESQ52F; coded for by C. elegans
cDNA yk134e9.5; coded for by C. elegans cDNA yk17b9.3;
coded for by C. elegans cDNA CEESK86B; coded for by C.
elegans cDNA yk17b9.3; coded for by C. elegans cDNA
yk134e9.3; coded for by C. elegans cDNA CEESQ40F"
/codon_start=1
/protein_id="AAB52260.1"
/db_xref="PIR:g1938422"
/db_xref="GI:1938422"
/translation="MAEPDPSAPVRLKTLTYNDGSPKKEPALSFTLIDSLICL
YDECSNLTREKCIAEFESVTVISKAKRLISRDPELVKYGKAGAEVAVWA
RGVEIYAMKILNKMEVRAETACREBHDVLYGRRMTNTHAFQGEANLYPM
DYIIGDMLTLLSKVDHIPSMAKYIAIMVLAIDHILGYHRRKRDVILNQ
GTRLDGSCRLRLADGSAVAVATPTIPEITIDWVSEEAQLRLQICSSD
VREGRNLSDFOLHPEEGIDWNTIRDSNPPYVEVSPEDTSNPDVDEDETPCE
TOPPRVLAATGMLPFVGFSTHGLSLSPALDEIRALACQCDALMEKSVG
FVLENERKAEVOKLEKQATIIAQAHAENPREDSENYSTINQDLDELINKRLE
DEVALOQOKKDEIVAESERKLEKELKRNKQMEKSEIORDINDLDIYVAK
ATVVOQDDMOAELADYDSLTLEKDSYKRLQDAEKAKQVAFBEKLETEREKA
LINKQEVITLARKSVETIDHLEEVAAKNTIASLATDEERTLEIKLKQORDER
ASHTAQSEQKQLEAEYERAKQMLQNVQMNVENRGLDEIEKLEKQMAALPRGL
NEQDLIEIFNVSEKATREMENTIRKIGEVSLKNSPLTSTNTQWPSGMSGR
RMNVARQDGLDOROLQAEIDAKTLKAEKNSQEQYLTSAAILDOTERMSIME
VAMLKQOKNIENSQSAFSTMGKMDIMNMNDEMNSIMQEKISQSTSVAN
AIIHHDQVFRKRDVLTAKPKMTAGTISPVSIAMERGHNERAKITPTICGHC

gene
CDS
/gene="K08B12.1"
join(29897..30145,30201..30626,30698..31147)
/gene="K08B12.1"
/note="similar to 11pases"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB52257.1"
/db_xref="PIR:g1938419"
/db_xref="GI:1938419"
/translation="MKREVFALAPATLALSQAGNVLVCKIDCKTCASYTHI
LGRFEDRMCLNETNGCPISCPIGRPVIOKDAFKPTPYKPAGRKRYDALRSIFA
VATAVDSNTECTLRNVRPDISRSYEVCDGSGNCKGILKSEBAKLYAYKGS
TAGROYAEMLHGTALQAGMEKESODAGVIVYFPAFRLEIDSGMEDDVIAMK
HNRVWLTGHSLGGLASATRALHLVKKYSVDNRKRLITRGPFRGNIAKAEIEN
IPRRYVRKRGDVPNNPAPLNAVILTAAYNQAMHYRLVHYDNNMGKGDFFNCS
VADLCGRNTRAMENLADHTSYFOVDDDFVRNCPGRLT"
BASE COUNT 11855 a 6960 g 12403 t
ORIGIN

Query Match 56.08; Score 22.4; DB 36; Length 37678;
Best Local Similarity 72.5%; Pred. No. 1.8e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 TGCCTAATCTCAATGCTGATTCATCAACAATGA 40
DB 709 TTCTGAATAATTATGTTGGAATTGATACGACGATGA 670

RESULT 13
AC006790/c
LOCUS 236496 bp DNA HTG 23-FEB-1999
DEFINITION Caenorhabditis elegans clone Y49F6B, WORKING DRAFT SEQUENCE, 1
unordered pieces.
ACCESSION AC006790
NID 64263148
VERSION AC006790.1 GI:4263148
KEYWORDS HTG; HTGS_Phasel.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.
1 (bases 1 to 236496)
REFERENCE 1
Waterson, R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 236496)
Waterson, R.H.
Direct Submission
Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 236496: contig of 236496 bp in length.
Location/Qualifiers

COMMENT
FEATURES

```

source
1. 236496
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y49F68"

BASE COUNT      73408 a 44340 c 44572 g 74176 t
ORIGIN

Query Match      55.5%; Score 22.2; DB 34; Length 236496;
Best Local Similarity 77.1%; Pred. No. 1.5e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      6 TAAATTCATAGCTGCTATTCATGAACAATGA 40
Db      54690 TAAATTCATAGCTATATTATTAAACAATGA 54656

RESULT 14
LOCUS      CELY49F6C 19566 bp DNA INV 06-FEB-1999
DEFINITION Caenorhabditis elegans cosmid Y49F6C.
ACCESSION AF125463
VERSION AF125463.1 GI:4226153
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditi; Rhabditiidae;
Rhabditiina; Rhabditiidae; Rhabditiidae; Pelodermatidae; Caenorhabditis.
1 (bases 1 to 19566)
Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Lalister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,
Saunders, D., Showkhen, R., Smalton, N., Smit, A., Sonnenmeier, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL
MEDLINE Nature 368 (6466), 32-38 (1994)
REFERENCE 94150718
AUTHORS 2 (bases 1 to 19566)
MILLER, N. and LE, T.
JOURNAL The sequence of C. elegans cosmid Y49F6C
REFERENCE Unpublished (1998)
AUTHORS 3 (bases 1 to 19566)
WATERSTON, R.
JOURNAL Direct Submission
TITLE Submitted (03-FEB-1999) Department of Genetics, Washington
JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
e-mail: rwnem@genome.wustl.edu and jess@sanger.ac.uk

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

FEATURES

source

This segment of YAC Y49F6 was sequenced to span the gap between F14D2 and T06D4. The 5' clone is F14D2, 200 bp overlap; 3' clone is T06D4, 200 bp overlap. Actual start of YAC Y49F6 is at base position 16839 of CELY14D2; actual end is at 14918 of CELY15D1.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFINDER (P. Green and L. Hillier, ms in preparation).

NEIGHBORING COSMID INFORMATION:

```

1. 19566
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/clone="Y49F6C"
/chromosome="II"
complement(1360..3975)
/gene="Y49F6C.6"
complement(join(1360..1499,1782..1985,2081..2179,
3387..3715,3863..3975))
/gene="Y49F6C.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAD12861.1"
/db_xref="PID:94226155"
/db_xref="GI:4226155"
/translation="MGITMGLTWTGVLATGATGATKISQHFKEFRNSQKVL
EHRFPFPISLKRESQEPNPSDPVLLQYISYTTVPPTPNNSKPSLOKX
TQFLENSNPVLLQYISYTTVPPTPNKLSISQOPFELSPRILVLEHER
CRGALSDNSVAVGAVCAKLVKTRHVKRISANSKLFPEFCYTYHTKNQV
SPERNANOKGKRMPEVYVALLFELNNVDIERSCSSSSFSRSLSSAYEYMAR
ALQILHNGVETIDREGGG"
complement(5272..6211)
/gene="Y49F6C.7"
complement(join(5272..5756,5891..5939,5988..6104,
6137..6211))
/gene="Y49F6C.7"
/codon_start=1
/evidence=not_experimental
/protein_id="AAD12861.1"
/db_xref="PID:94226161"
/db_xref="GI:4226161"
/translation="MPPEEHQNTLENYOPARTPVIFIPRNEIYAOIOVQDINROEI
RAREEDQRLRLDREMDNRIRIEMNIRAVLEENQSGSFSILFHSLSISLN
KSRPRTKEPKRPRKPKRQKRTKPALEFDGLAHTCLKGLSLALMLSLFSE
YRPGKRAHRAQTTHAIIFFSVNDQRLVHEGATLEPVHRSASFENLFFQOLL
YAFLOFSDNGSRAPSRNRNAGG"
complement(7296..8106)
/gene="Y49F6C.8"
complement(join(7296..7577,7848..7979,8029..8106))
/gene="Y49F6C.8"
/codon_start=1
/evidence=not_experimental
/protein_id="AAD12860.1"
/db_xref="PID:94226154"
/db_xref="GI:4226154"
/translation="MAQPNFSFTSFGMTINMLVSTRAGLEFELTIKOPTYQNSIQRI
QEEHREKQYLROALIDLADSEEDGQPLDEKTSKMAELTIENNNIRSFNIV
NEORTGSKDROHMEERIQHWTKEDREELSGLAVMKOTIOESMKDAADDERKSR
NNO"
9324..10159
/gene="Y49F6C.5"
join(9324..9725,9785..9932,9978..10159)
/gene="Y49F6C.5"
/notes="contains similarity to BTB (also known as BR-C/TLX)
domains (Pfam:PF00651, Score=58.8, E=1.2e-13, N=1)"
/codon_start=1
/evidence=not_experimental

```

Query Match	55.5%	Score 22.2	DB 37	Length 19566
Best Local Similarity	77.1%	Prod. No. 2.3e+02		
Matches 27	Conservative	0	Mismatches 8	Indels 0
QY	6	TAATATCTAATGCTGATTTTCATGAACAATGA	40	
Db	14055	TAATATCAATTAATTAATTTATTTAAACAATATA	14021	
RESULT 15				
LOCUS	A37232	3498 bp	DNA	
DEFINITION	Sequence 1 from Patent WO9403609.		PAT	05-MAR-1997
ACCESSION	A37232			
NID	92294345			
VERSION	A37232.1	GI:2294345		
KEYWORDS				
SOURCE	unidentified.			
ORGANISM	unidentified			
REFERENCE	1 (bases 1 to 3498)			
AUTHORS	Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.			
TITLE	EUCARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE OR PROTEIN KINASE AND ASSAYS USING THEM			
JOURNAL	Patent: WO 9403609-A 1 17-FEB-1994;			
COMMENT	IMP CANCER RES TECH (GB)			
FEATURES	Other publication JP 85031247 960409.			
source	location/Qualifiers			
	1..3498			
	/organism="unidentified"			
	/db_xref="taxon:32644"			
	1..3207			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="CA02284.1"			
	/db_xref="PID:g304808"			
	/db_xref="PID:g2294346"			
	/db_xref="GI:2294346"			
	/translation="MPRSSGGLGIVGILHMPRLVIECLLPNGMIVTLECLREALLITIKHELEAKRKYPLHOLLDESSYIVSVTOGEEREFEPETRLDRLPOPLKVIYISERGNKEKILNRELIFATGMPCFEDMVDPEVQDFRNIILVCKEAYDLRLNSPHSRMAYIPRVESPELPKPHITINKLIDSGQIIVIVIVSGVNNRQKTLIKINHCVPEQVAEAIIRKTRSMLLSEQLKCVLEIOGQYILKVCQCEVLEKPTLSQYXIRSGALIMIGRPNNMLMAKESILYSQLPDCFTMPSSRSRISTATPNNGENSTKSLAYNSALRIITLCATYVNVNIRIDIKYRTGTYHGGECDLVNVTORVCSNPRNMENIYDIYJDLPLPRAARILSICSVKGRKAKKEBCHPLAMGNINFDYDPLVSKMALNIMPVPRGLEDLNPIGVTCISNPKETPCLELEEDWFSSVYKFPDMKSVIEBHAKSVSREGESYSNAGLSNRLARDNELRENDKEBQRLAICTRDLSEITIQEKDFLWSHHYCVTVTEILPDKLSTSVNSRDEVAQMYCLVMDMPRIKPEQAAELLDCNYPDMVNRGFAVACLEKTYQDLSYQLIOLVOVLKYEQYLDNLVAFKLKALTNORIGHPEFMHLSMHNKTVSQARFLLLESCACRACGMILKHLNRQVAMEKILINTDLIKOKRDETOKVOKPLVEQMRARPDPMALOGFLSPINPAHDLGKRLIECCINSARKPLIMNENPDIVSELPQONNEIIFKNDDLEKQMDLTQIIRIKEMINOMQGLDLMPLYGCLSGDVGLELVNRNHTIMQIOQKGLKALQFNSTHLMQLDKNNKGEIYDAIDFTSSCAQCVATFLJGIGDRNINSINWKDCQLFHDHGHFLDHRKKKGRKREVPVLVQDFEIVLSKGOECTTKTEEFPEQEMCYKAYIALIRQHANLFLNLESXMGSGMPELQSFDDIAYIRKTLADKTDDEALEYMKQNDAAHGGCTTKMGMIFHTIKQHANL"			
BASE COUNT	1147 a	626 c	722 g	1003 t
ORIGIN				
Query Match	55.0%	Score 22	DB 5	Length 3498
Best Local Similarity	73.7%	Prod. No. 3.5e+02		
Matches 28	Conservative	0	Mismatches 10	Indels 0
QY	1	TGCTGTAATTCATATGCTGATTTTCATGAACAAT	38	
Db	3325	TGCTGTAATTCATATGCTGATTTTCATGAACAAT	3288	

Search completed: September 13, 1999, 16:19:58

Job time: 4510 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:21:51 ; Search time 133.08 Seconds

(without alignments)
75.201 Million cell updates/sec

Title: US-09-325-095-2122.SEQ

Perfect score: 40

Sequence: 1 TGCCTGAATTCCTAATGCTGATTCATGAACAATGA 40

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	55.0	3498	1 057012	Ptdins 3-kinase 11
2	22	55.0	1440	1 V74847	Staphylococcus aur
3	21.6	54.0	110000	1 V21209_11	Continuation (12 o
4	21.6	54.0	110000	1 V21209_12	Continuation (13 o
5	21.4	53.5	652	1 V74927	Staphylococcus aur
6	21.2	53.0	1192	1 T08635	Hpv E6-binding pro
7	21	52.5	110000	1 T58840_0	Mycoplasma genital
8	21	52.5	184	1 V11694	Divericin A DNA.
9	21	52.5	1431	1 V53139	Helicobacter seryl
10	21	52.0	593	1 X17718	Divericin A signa
11	20.8	52.0	2859	1 085373	Growth factor-acti
12	20.8	52.0	110000	1 T28960	Human 26S proteaso
13	20.8	52.0	110000	1 T58840_1	Continuation (2 of
14	20.8	52.0	786	1 V28681	Ripening banana pu
15	20.6	51.5	4810	1 V70887	Plasmid pTK-fus-1a
16	20.4	51.0	1590	1 V27169	Bacillus thuringie
17	20.2	50.5	288	1 V88567	EST clone FK295. N
18	20	50.0	3412	1 051156	Human p110 CDNA. R
19	20	50.0	3207	1 051155	p110 CDNA. Recombi
20	20	50.0	20	1 059020	Antisense PCR prim
21	20	50.0	20	1 059021	Sense PCR primer f
22	20	50.0	3498	1 057012	Ptdins 3-kinase 11
23	20	50.0	2766	1 T75488	DNA for Bacillus s
24	20	50.0	192	1 V88282	EST clone GR410. N
25	20	50.0	3501	1 X13430	Enterococcus faeca
26	20	50.0	9767	1 X13067	Enterococcus faeca
27	20	50.0	725	1 X13997	H. pylori GHP 876
28	20	50.0	110000	1 X20248_06	Continuation (7 of
29	20	50.0	111309	1 X20250	Borrelia burgdorfe
30	19.8	49.5	3932	1 064132	cytR5 gene. Isola
31	19.8	49.5	2288	1 083526	P. falciparum EB1-
32	19.8	49.5	4948	1 T42134	12C-1 gene encodin
33	19.8	49.5	3934	1 T68434	H. pylori secreted
34	19.8	49.5	873	1 T68271	Plasmid encoding ebl-1 g
35	19.8	49.5	2288	1 T72894	DNA encoding Bacil
36	19.8	49.5	3934	1 T95081	MSRV-1 clone LTRGA
37	19.8	49.5	2503	1 T96475	Clone LTRGAG12 DNA
38	19.8	49.5	2503	1 V43169	DNA encoding a Cry
39	19.8	49.5	3934	1 V83927	H. pylori secreted
40	19.8	49.5	870	1 X30444	Type 5 17-beta-hyd
41	19.6	49.0	700	1 T63475	Novel protein from
42	19.6	49.0	2095	1 V12251	Enterococcus faeca
43	19.6	49.0	17764	1 X13238	

44 19.6 49.0 11879 1 X13118
45 19.4 48.5 1262 1 063213

Enterococcus faeca
Transcription acti

ALIGNMENTS

```

RESULT 1
ID 057012/c
AC 057012:
DT 31-AUG-1994 (first entry)
DE Ptdins 3-kinase 110 kD catalytic subunit cDNA.
KW 110 kD catalytic subunit; phosphatidyl inositol 3-kinase;
KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
KW blood vessel plaques; ss.
OS Bos taurus.
FH Key
FT cds
FT 1..3207
FT /*tag= a
FT /*product= p110
PD 17-FEB-1994.
PR 05-AUG-1993; G01651.
PR 05-AUG-1992; GB-016654.
PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
DR WPI: 94-065697/08.
DR P-PSDB: R46294.
PT Eukaryotic cells transformed with mammalian phospholipid or
PT protein kinase DNA - useful in assays for compounds involved in
PT cell growth regulation and for treating cancers
PS Disclosure: Fig 1; 71pp; English.
CC This sequence encodes the 110 kD catalytic subunit of the phosphatidyl
CC inositol (Ptdins) 3-kinase. This sequence was transformed into
CC Schizosaccharomyces pombe cells under the regulatory control of the
CC nmt promoter in an embodiment of the invention. In the presence of
CC thiamine the promoter is inactive and the cells carrying the Ptdins
CC catalytic subunit plasmid grow as the parental strain. In the absence
CC of thiamine the nmt promoter functions and the Ptdins 3-kinase
CC catalytic subunit is induced. Ptdins activity is substantially
CC increased under these conditions. Cells containing constructs such as
CC this, are useful in assays for detecting compounds involved in cell
CC growth regulation. It is also used as the basis for detecting
CC compounds for treating cancers and the formation of blood vessel
CC plaques.
SQ Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 55.0%; Score 22; DB 1; Length 3498;
Best Local Similarity 73.7%; Pred. No. 17;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 1 TGCCTGAATTCCTAATGCTGATTCATGAACAAT 38
DB 3325 TGCCTGAATTCCTAATGCTGATTCATGAACAAT 3288

RESULT 2
ID 057012/c
AC 057012:
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #536.
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key
FT misc_feature
FT 121..180
FT /*tag= a

```



```

FT      encoded protein shows 26.82 percentage
FT      identity to ATP-dependent nuclease (addn)
FT      from B. subtilis"
FT      cds
FT      35242..35904
FT      /*tag- j
FT      /label- MG033
FT      /note- "Previously identified as MORF-20100, the
FT      encoded protein shows 35.90 percentage
FT      identity to glycerol uptake facilitator
FT      (glpF) from B subtilis"
FT      complement (39873..40514)
FT      /*tag- x
FT      /label- MG034
FT      /note- "Previously identified as MORF-20101, the
FT      encoded protein shows 48.13 percentage
FT      identity to thymidylate kinase (tdk)
FT      from B. subtilis"
FT      cds
FT      40543..41787
FT      /*tag- l
FT      /label- MG035
FT      /note- "Previously identified as MORF-20102, the
FT      encoded protein shows 30.71 percentage
FT      identity to histidyl-tRNA synthetase (hls5)
FT      from Mycobacterium leprae"
FT      complement (44751..46277)
FT      /*tag- m
FT      /label- MG038
FT      /note- "Previously identified as MORF-20105, the
FT      encoded protein shows 46.89 percentage
FT      identity to glycerol kinase (glpk)
FT      from E. coli"
FT      complement (46268..47422)
FT      /*tag- n
FT      /label- MG039
FT      /note- "Previously identified as MORF-19831 and
FT      MORF-20106, the encoded protein shows 43.20
FT      percentage identity to glycerol-3-phosphate
FT      dehydrogenase (GUT2) from S. cerevisiae"
FT      49377..49643
FT      /*tag- o
FT      /label- MG041
FT      /note- "The encoded protein shows 48.86 percentage
FT      identity to phosphotransferase (ptsh) from Mycoplasma
FT      capricolum"
FT      cds
FT      50060..51520
FT      /*tag- p
FT      /label- MG042
FT      /note- "Previously identified as MORF-19832 and
FT      MORF-20108, the encoded protein shows 41.92
FT      percentage identity to spermidine/
FT      putrescine transport ATP-binding protein
FT      (potA) from E. coli"
FT      51525..52382
FT      /*tag- q
FT      /label- MG043
FT      /note- "Previously identified as MORF-20110, the
FT      encoded protein shows 26.51 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein (potB) from E. coli"
FT      52366..53220
FT      /*tag- r
FT      /label- MG044
FT      /note- "Previously identified as MORF-20111, the
FT      encoded protein shows 29.45 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein C (potC) from E. coli"
FT      54658..55605
FT      /*tag- s
FT      /label- MG046
FT      /note- "Previously identified as MORF-20112, the
FT      encoded protein shows 36.60 percentage
FT      identity to sialoglycoprotease (gcp)

```

Query Match	Best Local Similarity	Matches	Score	DB	Length
52.5%	73.0%	27	21	1	110000
Conservative	0	Mismatches	10	Indels	0
Gaps	0				
Query	4	TTTTAATTCGTAAGCTGTTTCAATCAATAAATA	20891		
DB	20927	TTTTAATTCGTAAGCTGTTTCAATCAATAAATA	20891		
RESULT	8				
ID	V11694				
AC	V11694				
DT	14-SEP-1998				
DE	Divergicin A DNA.				
FW	Divergicin A: bacteriocin; signal peptide; protein secretion; preservation; food spoilage; lactic acid bacterium; antibacterial;				
SS	SS.				
OS	Carnobacterium divergens.				
FH	Key	Location/Qualifiers			
FT	RBS	24..28			
FT		/*tag- a			
FT	sig_peptide	38..125			

```

FI      mat_peptide      /tag- b
FT      126..184
TI      /tag- c
TI      114..144
PI      primer_bind      /tag- d
PI      WC0909639-A1.
PD      12-MAR-1998.
PF      05-SEP-1997; U15609.
PR      05-SEP-1996; US-026257.
PA      (UNAL-) UNIV ALBERTA.
PI      Franz CM, Greer GG, Leisner JT, McCormick JK, McMullen LM,
PI      Poono A, Stiles ME, Van Belkum MJ, Vederas JC,
PI      Worobo RJ, Morobo RW;
PI      WPI: 98-193319/17.
DR      P-PSDB; W59007.
PT      Bacterial growth inhibiting peptide(s) enterocin 900 or brochocin C
PT      - used to inhibit growth of susceptible bacteria in e.g. foodstuff,
PT      laying animal, food preparation area or fermentation vessel
PS      Example 2: Page 115; 174pp; English.
CC      This DNA sequence codes for divergicin A (see W59007), a bacteriocin
CC      produced by the meat lactic acid bacterium, Carnobacterium divergens.
CC      The invention provides a secretion vector in which the structural and
CC      immunity genes of other bacteriocins, such as carnobacteriocin B2
CC      (see W59029-31), colicin V, leuocin A and brochocin-C (see W59021),
CC      are fused to the divergicin A signal peptide (see W59010) gene
CC      sequence, allowing production and secretion of the bacteriocin(s) by
CC      host and heterologous bacteria. Bacteriocin cassettes encoding the
CC      divergicin A signal peptide and two or more bacteriocins are used to
CC      tailor the antibacterial spectrum of a producer strain to target a
CC      range of spoilage or pathogenic bacteria, e.g. Escherichia coli and
CC      Salmonella. Using the technology of multiple bacteriocin production
CC      and delivery using lactic acid bacteria, a range of bacteriocins can
CC      be produced in situ in human food, animal feed, the gastrointestinal
CC      tract of humans and animals, and in the female genital tract.
SQ      Sequence 184 BP; 60 A; 24 C; 37 G; 63 T;

```

```

Query Match      52.5%; Score 21; DB 1; Length 184;
Best Local Similarity 73.0%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

```

OY      1 TCGTAAATCTATGCTGATTTTCATGAAACAA 37
        |||||
DB      121 TCGTAAATCTATGCTGATTTTCATGAAACAA 157

```

```

RESULT 9
V53139      standard; DNA: 1431 BP.
AC      V53139;
DT      12-NOV-1998 (first entry)
DE      Helicobacter seryl-tRNA synthetase encoding DNA.
KW      Helicobacter; aminoacyl-tRNA synthetase; isoleucyl-tRNA synthetase;
KW      methionyl-tRNA synthetase; leucyl-tRNA synthetase; valyl-tRNA synthetase;
KW      lysyl-tRNA synthetase; seryl-tRNA synthetase; human gastric mucosa;
KW      chronic gastritis; antimicrobial compound; antisense inhibition;
KW      H pylori; bacterium; enzyme; ds.
OS      Helicobacter pylori.
FH      Key      Location/Qualifiers
FT      CDS      80..1327
FT      /tag- a
FT      /product= "seryl-tRNA synthetase"
FT      US5801013-A.
PD      01-SEP-1998.
PF      26-MAY-1995; 451715.
PR      26-MAY-1995; US-451715.
PA      (CUBI-) CUBIST PHARM INC.
PI      Houman F, Oiu Y, Schimmel PR, Shen X, Tao J;
PI      WPI: 98-494769/42.
DR      P-PSDB; W75914.
PT      DNA encoding Helicobacter amino-acyl-tRNA synthetase proteins -
PT      useful for recombinant production of the enzyme, anti-sense
PT      constructs and hybridisation probes, and construction of tester

```

```

PT      strains to test for inhibitors of the enzyme
PS      Claim 5; Columns 95-100; 76pp; English.
CC      This DNA encodes a Helicobacter seryl-tRNA synthetase. The invention
CC      provides nucleic acid sequences (V53134 to V53139) encoding Helicobacter
CC      aminoacyl-tRNA synthetases (hARNAs) (W75909 to W75914), where the
CC      aminoacyl-tRNA synthetase is isoleucyl-tRNA synthetase, methionyl-tRNA
CC      synthetase, leucyl-tRNA synthetase, valyl-tRNA synthetase, lysyl-tRNA
CC      synthetase or seryl-tRNA synthetase respectively and at least a portion
CC      of the hARNAs has catalytic activity or binding function. The nucleic
CC      acids and recombinant cells containing the nucleic acids can be used for
CC      recombinant production of the enzymes. The nucleic acids can also be used
CC      as hybridisation probes in experiments to identify hARNAs sequences.
CC      H. pylori is bacteria that infects human gastric mucosa, leading to
CC      chronic gastritis. The amino acyl tRNA synthetases are used by the
CC      bacterium during protein synthesis, and as such can be a target for
CC      antimicrobial compounds. The tester strains can be used to screen for
CC      such compounds, and antisense constructs based on the nucleic acids can
CC      be used for antisense inhibition of the enzyme. The proteins themselves
CC      can be used to raise antibodies, which in turn can be used for
CC      purification and study of the enzyme. Screening of inhibitors of the
CC      enzyme, as well as antisense antimicrobial activity through the nucleic
CC      acids are novel approaches to inhibition activity against the bacteria,
CC      as many present antibiotics used have side-effects and the bacteria are
CC      becoming resistant to them.
SQ      Sequence 1431 BP; 498 A; 271 C; 320 G; 342 T;

```

```

Query Match      52.5%; Score 21; DB 1; Length 1431;
Best Local Similarity 82.8%; Pred. No. 35;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

OY      11 TCTAATGCTGATTTTCATGAAACAAATG 39
        |||||
DB      619 TTTAATGCTGATTTTCATGAAACAAATG 647

```

```

RESULT 10
X17718      standard; DNA: 184 BP.
ID      X17718;
DT      25-MAY-1999 (first entry)
DE      Divergicin A signal peptide encoding DNA.
KW      Enterocin 900; Brochocin-C; bacterial growth; microorganism; inhibit;
KW      food; meat; fermentation; bacteriocin; mastitis; secretory vector; lacc;
KW      leuocin A; lacc; enzyme; probiotic; L. gelidum; Listeria; Lactobacillus;
KW      sulphide-producing; spoilage; sensory property; smell; taste; infection;
KW      ground beef; silage; Salmonella; poultry; intestine; therapeutic; colour;
KW      mucosal tissue; vacuum-packed; storage; divergicin A; ss.
OS      Carnobacterium divergens.
FH      Key      Location/Qualifiers
FT      RBS      24..28
FT      /tag- a
FT      38..184
FT      CDS      38..125
FT      /tag- b
FT      114..144
FT      primer_bind /tag- d
PI      WC0902555-A1.
PD      21-JAN-1999.
PF      09-JUL-1998; U14331.
PR      05-SEP-1997; US-924629.
PA      (UNAL-) UNIV ALBERTA.
PI      Franz C, Greer GG, Leisner JT, McCormick JK, McMullen LM,
PI      Poono A, Stiles ME, Van Belkum MJ, Vederas JC, Worobo RJ,
PI      Worobo RW;
PI      WPI: 99-120779/10.
DR      P-PSDB; W94865.
PT      New bacteriocins enterocin 900 and brochocin-C - useful as food
PT      preservatives and therapeutic antimicrobials
PS      Disclosure: Page 115; 182pp; English.
CC      The invention relates to peptides enterocin 900, brochocin-C that inhibit

```

CC bacterial growth. Microorganisms expressing the above peptides are used
 CC to inhibit bacterial growth in foods (specifically meat), live animals
 CC (applied topically), food preparation areas and fermentation vessels. The
 CC bacteriocin enterocin 900, brochoxin-C may be used directly, specifically
 CC to inhibit bacteria that cause mastitis. Antibodies specific for the
 CC peptides are used to detect expression of bacteriocins in cells. Secretory
 CC vectors (containing sequences for a promoter, a leucocin A processing
 CC peptide, a heterologous peptide, leac and leac) are used to secrete
 CC bacteriocins, enzymes or other proteins, e.g. for use in food production
 CC and as probiotics. L. gelidum, or other leucocin A-expressing bacteria,
 CC inhibit the growth of *Listeria* and the sulphide-producing spoilage
 CC organism *Lactobacillus sakei*; they also improve the sensory properties
 CC (smell, taste and colour) of meat, specifically ground beef. L. gelidum
 CC can also be used in preparation of animal feeds (silage), as probiotic,
 CC to control *Salmonella* in poultry intestines and therapeutically against
 CC mucosal tissue infections. Treatment with bacteriocins provides vacuum-
 CC packed meat products with predictable and longer storage life. The
 CC secretion vectors can express several bacteriocins for broader spectrum
 CC of activity, e.g. tailored for a particular target organism.
 SQ Sequence 184 BP; 60 A; 24 C; 37 G; 63 T;

Query Match

Best Local Similarity 52.0%; Score 21; DB 1; Length 184;
 Pred. No. 29;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 1 TGCTGAATTCGATGCTGATTCGAAACAAGA 37
 |||||
 DB 121 TGCTGAATTCGATGCTGATTCGAAACAAGA 157

RESULT 11

085373
 ID 085373 standard; cDNA: 593 BP.

DT 29-AUG-1995 (first entry)
 DE Growth factor: fic gene; heparanase; heparin; heparan sulfate;
 KM Growth factor: fic gene; heparanase; heparin; heparan sulfate;
 OS Homo sapiens.
 PN WO9504158-A.
 PD 09-FEB-1995.
 PF 26-JUL-1994; U08207.
 PR 29-JUL-1993; US-099866.
 PR 13-OCT-1993; US-136117.
 PA (UPJO) UPJOHN CO.
 PI Hoogwerf AJ, Ledbetter SR;
 DR WPI: 95-082239/11.
 DR P-PSDB: R70803.
 PT Screening for cpds. with anti-heparanase activity - by detecting
 PT inhibition of heparin or heparan sulphate degradation, cancer.
 PT potentially useful for treating arthritis, restenosis, cancer.
 PS Disclosure; Page 53; 60pp; English.
 CC Purified heparanases, prepared under reducing conditions and
 CC activated with trisglutaminase, are given in R70786-805. Most
 CC are prepared by reverse transcription of mRNA from activated human
 CC leukocytes, then cloning of the cDNA into pVIL1392 Daculovirus
 CC vector, and expression in 579 cells in the presence of reduced
 CC glutathione and dithiothreitol.
 SQ Sequence 593 BP; 170 A; 136 C; 135 G; 152 T;

Query Match

Best Local Similarity 52.0%; Score 20.8; DB 1; Length 593;
 Pred. No. 38;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 9 AATTCATGCTGATTCATGAACAAGA 40
 |||||
 DB 124 ATCCACATGCTGATTCATGAACAAGA 155

RESULT 12

128960/c

ID 128960 standard; cDNA to mRNA; 2859 BP.

AC 128960;
 DT 28-NOV-1996 (first entry)
 DE Human 26S proteasome constitutive protein p12 cDNA.
 KW Human; 26S proteasome; constitutive protein; p12; treatment;
 OS Homo sapiens.
 FT Key Location/Qualifiers
 FT cds 1..2859
 FT /*tag= a
 FT /note= "STOP codon absent"
 PN J08116972-A.
 PD 14-MAY-1996.
 PF 28-OCT-1994; 264810.
 PR 28-OCT-1994; JP-264810.
 PA (KYOM) KYOMA HAKKO KOGYO KK.
 PI Tanaka K, Yamasaki M, Yano K;
 DR WPI: 96-280771/29.
 DR P-PSDB: R97284.
 PT Human 26S proteasome constitutive proteins p12 and p21 - useful to
 PT diagnose and treat conditions associated with human 26S proteasome
 PT related diseases, e.g. Alzheimer's
 PS Claim 3; Pages 9-10; 12pp; Japanese.
 CC The present sequence encodes the human 26S proteasome constitutive
 CC protein p12, which can be used to treat and diagnose 26S
 CC proteasome associated conditions, e.g. Alzheimer's disease, and in
 CC the study and elucidation of human 26S proteasome function. A cDNA
 CC fragment was synthesised by RT-PCR (science 241, 708-712 (1988)),
 CC on the basis of an amino acid sequence corresponding to the bovine
 CC proteasome constitutive protein p12. The cDNA fragment was used
 CC as a plaque hybridisation probe to isolate phages contg. human
 CC p12 protein DNA, isolated from a human hepatic, or HepG cell line
 CC (ATCC HB8065) cDNA library. The phages were introduced into hosts
 CC by the in vitro packaging method. From the resulting transformants
 CC the protein was isolated and purified.
 SQ Sequence 2859 BP; 875 A; 558 C; 672 G; 754 T;

Query Match

Best Local Similarity 52.0%; Score 20.8; DB 1; Length 2859;
 Pred. No. 44;
 Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 1 TGCTGAATTCGATGCTGATTCGAAACAAGA 40
 |||||
 DB 1272 TGCCATTAACTGATGCTCTTTTCATGACCTTATGA 1233

RESULT 13

T58840.1/c
 Continuation (2 of 6) of T58840 from base 100001 (Mycoplasma genitalium genome.)
 WP Sequence split into 6 fragments LOCUS T58840 Accession T58840
 WP Fragment Name Begin End
 WP T58840_0 1 110000
 WP T58840_1 100001 210000
 WP T58840_2 200001 310000
 WP T58840_3 300001 410000
 WP T58840_4 400001 510000
 WP T58840_5 500001 580073

Query Match

Best Local Similarity 52.0%; Score 20.8; DB 1; Length 110000;
 Pred. No. 60;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 7 AATTCATGCTGATTCATGAACAAGA 38
 |||||
 DB 33361 AATTCCAAGCTAGTAATTCGTAATAAT 33330

RESULT 14

V28681/c
 ID V28681 standard; cDNA: 786 BP.
 AC V28681;

CC recognised by vertebrate transcription factors and operatively
CC linked to a heterologous gene, which allows expression of the
CC heterologous gene when infected into a vertebrate cell. The rBPV

Search completed: September 13, 1999, 16:22:01
Job time: 4021 sec

THIS PAGE BLANK (uspto)

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

FEATURES

SOURCE

High quality sequence stops: 169 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1713 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 169.
Location/Qualifiers

1..231
/organism="Homo sapiens"
/db_xref="GDB:401930"
/db_xref="LXON:9606"
/clone="IMAGE:29583"
/clone_lib="Soares Infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: latmid BA; Site: 1: Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5';
ACCTGAGAGATTCGCGCCGACAGCAATTTTATTTTATTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the latmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 94 a 24 c 29 g 83 t 1 others
ORIGIN

Query Match

Best Local Similarity 59.0%; Score 23.6; DB 22; Length 231;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 CTGTAAATCTATGCTGATTCATGAAACCAATGA 40
||||| ||||| ||||| ||||| ||||| |||||
Db 14 CTGGAATAATATAATATTTTATTCATAAAGAAATGA 51

RESULT 4 AA936397 361 bp mRNA EST 24-AUG-1998
LOCUS AA936397.1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1371317 3',
DEFINITION mRNA sequence.
ACCESSION AA936397
NID 93094315
VERSION AA936397.1 GI:3094315
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
AUTHORS Eutheria: Primates: Catarrhini: Homiidae: Homo.
1 (bases 1 to 361)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
REFERENCE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 17, 1998 this sequence version replaced gi:2044601.

JOURNAL COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/db/brp/image/image.html

FEATURES

SOURCE

Insert Length: 985 Std Error: 0.00
Seq primer: -40ml3 fwd. RT from Amersham
High quality sequence stop: 333.
Location/Qualifiers

1..361
/organism="Homo sapiens"
/db_xref="LXON:9606"
/clone="IMAGE:1371317"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site: 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGACGAGGAGCGCCGCCCAATTTTATTTTATTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 98 a 68 c 92 g 103 t
ORIGIN

Query Match

Best Local Similarity 56.0%; Score 22.4; DB 40; Length 361;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 9 ATTCTAATGCTGATTCATGCAAAACAAATGA 40
||||| ||||| ||||| ||||| ||||| |||||
Db 297 ATTCTAATGCAAGATATTCATGATCCATCA 266

RESULT 5 AV047057 305 bp mRNA EST 18-MAY-1999
LOCUS AV047057 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone
DEFINITION 170065SH12, mRNA sequence.
ACCESSION AV047057
NID 94866722
VERSION AV047057.2 GI:4866722
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
AUTHORS Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 305)
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Ito, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,
Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H.,
Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,
Sugahara, Y., Suzuki, H., Suzuki, R., Tateo, M., Tomaru, Y.,
Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
On Jun 22, 1998 this sequence version replaced gi:3247212.

JOURNAL COMMENT
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098


```

OY      1  TGCTGAATTTCTATGCTGCTATTCATGAACA  35
        || || | ||| ||||| ||| |||
DB      191 TGATGAATTTGTAATACTGCTATTCATGACAGA  155

```

RESULT
AA866471

Accession	LOCUS	DEFINITION	EST
	A88664.71	554 bp mRNA	05-FEB-1999
	U1-R-E0-br-9-08-0-U1.s1	Rattus norvegicus CDNA clone	
	U1-R-E0-br-9-08-0-U1.3'	similar to gb AA123324 AA123324.mn37f10.r	
	Beddinghouse mouse embryonic region Mus musculus CDNA clone 540139		
	5', mRNA sequence.		

ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANIS

REFERENCE
1 (bases 1 to 554)
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REVISIONS	
TITLE	Normalisation and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	On Mar 16, 1998 this sequence version replaced g1:2961932.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

The sequencing present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult 18-Day-Embryo library. cDNA Library Preparation: M. Bettina Bonaldi, Ph.D. Clone distribution: Clones will be available through Research Genetics. The following repetitive elements were found in this cDNA sequence: 103-170, >AT-rich#low_complexity leg primer: M13 Forward.

FEATURES	Location/Qualifiers
source	1. .554

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_uid="R-E0-Dt-g-08-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pUT7SD-Pac (Pharmacia) with a modified
polylinker Site_1: NotI; Site_2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."

```

BASE COUNT	187 a	89 c	83 g	195 t
ORIGIN				

Query Match	55.58;	Score 22.2;	DB 45;	Length 554;
Best Local Similarity	77.18;	Pred. No. 2.2e+02;		
Matches 27; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0

QY 1 TGCCTGAATCTCTAATGCTGTAATTCGAACA 35
|||||
Db 211 TGCCTTAATGCTAATGCAACATTTCTCGAACA 245

RESULT 9

LOCUS	475 bp	mRNA	EST	13-FEB-1996
DEFINITION	YY39a10.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone			

ACCESSION	N44819	
NID	g1185985	
VERSION	N44819.1	GI:1185985
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia
Eutheria; Primates; Catarrhini; Hominae; Homo.
1 (bases 1 to 475)
Hiller, L., Clark, N., Dubucque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kuebb, T., Le, M., Lennon, G., Marra, N.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., and
Wilson, R.

TITLE	The WashU-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	On May 8, 1995 this sequence version replaced gi:800198

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Tel: 314 286 1800
Fax: 314 286 1810

Email: estewartson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMaCS Consortium (info@image.llnl.gov) for further information.
Seq primer: T7

High quality sequence stop: 402

FEATURES
SOURCE

1. .475
/organism="Homo sapiens"
/db_xref="GDB:3883236"
/db_xref="taxon:9606"
/clone="IMAGE:273594"
/clone_1lb="Scates melanocyte 2nbhm"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker. Site_1: Not I. Site_2: Eco RI. 1st strand cDNA
was primed with a Not I - CCGAG(GT) primer [5].
TGTTACCAATCTGAAGGAGGGCGGCGGATTTTTTTTTTTT 3'}.
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Felicia Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."
90 c 85 g 128 t
2 others

Query Match	55.0%	Score 22:	DB 25:	Length 475:
Best Local Similarity	73.7%	Pred. NO.	2.7e+02:	
Matches 28:	Conservative	0:	Mismatches 10:	Indels 0:
				Gaps 0

Oy	1	TGCTGTAAATTCATATCGTGATTCCATGAACCAAT	38
Db	372	TGCTGTAAATTCATATCGTGATTCCATGAACCAAT	335

RESULT	10	
AA298585/c		
LOCUS		
AA298585	343 bp	mRNA
EST114216	HSC172 cells	II Homo sapiens
DEFINITION		

AA296585	343 bp	MRNA	EST	18-APR-1997
EST144216	HSC172 cells	II Homo sapiens	cDNA 5' end	similar to
phosphatidylinositol 3-kinase,				
alpha catalytic subunit, mRNA.				

ACCESSION AA298585
 MID g1950938
 VERSION AA298585.1 GI:1950938
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 343)
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Melnick,D.K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Batle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghagen,N.S., Glodde,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R., Small,R.V., Spriggs,T.A., Utechtack,T.R., Weisman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepede,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,T., Li,H., Mellesner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Pannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M., and Venter,J.C.
 Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl), 3-174 (1995)
 96026280
 On Sep 12, 1996 this sequence version replaced g1:1392803.
 MEDLINE
 COMMENT Other_ESTs: TRC168479
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavage@tigr.org
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..343
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):179595"
 /db_xref="taxon:9606"
 /clone_lib="HSCI172 cells II"
 /cell_type="fibroblast"
 /dev_stage="fetal"
 /note="Organ: Lung; Vector: pbluescript SK-; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 132 a 61 c 64 g 84 t 2 others
 ORIGIN
 Query Match 55.0%; Score 22; DB 31; Length 343;
 Best Local Similarity 73.7%; Pred. No. 3e+02;
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 Oy 1 TCGTGAATTCATGCTGATTCATGCAACAAT 38
 ||||||||||||||| | | | |
 Db 287 TCGTGAATTCATGCTGATTCATGCAAT 250
 RESULT 11
 LOCUS A1632172 633 bp mRNA EST 26-APR-1999
 DEFINITION ts85507.x1 NCI-CGAP_G6 Homo sapiens cDNA IMAGE:2238108 3' similar to SW:PIIA.BOVIN P32871 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;, mRNA sequence.

ACCESSION A1632172
 MID g4683502
 VERSION A1632172.1 GI:4683502
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 633)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 On May 7, 1998 this sequence version replaced g1:3121411.
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskajuk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLNT at: www-bio.llnl.gov/bdnp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 446.
 Location/Qualifiers
 1..633
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="15g26.1"
 /clone_lib="IMAGE:2238108"
 /clone_lib="NCI-CGAP_G6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP_G64 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo.
 BASE COUNT 169 a 109 c 108 g 246 t 1 others
 ORIGIN
 Query Match 55.0%; Score 22; DB 49; Length 633;
 Best Local Similarity 73.7%; Pred. No. 2.5e+02;
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 Oy 1 TCGTGAATTCATGCTGATTCATGCAACAAT 38
 ||||||||||||||| | | | |
 Db 220 TCGTGAATTCATGCTGATTCATGCAAT 257
 RESULT 12
 LOCUS AV024880 246 bp mRNA EST 02-JUN-1999
 DEFINITION AV024880 Mus musculus adult C57BL/6J lung Mus musculus cDNA clone 1200008805, mRNA sequence.
 ACCESSION AV024880
 MID 94801872
 VERSION AV024880.1 GI:4801872
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE
AUTHORS

Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 246)
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Ito, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomioka, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
On Mar 16, 1998 this sequence version replaced gi:2961906.

TITLE
JOURNAL
COMMENT

Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rken.go.jp
Thermolabile and thermostable activation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

source

1. 246
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1200008B05"
/clone_1lb="Mus musculus adult C57BL/6J lung"
/tissue="lung"
/dev_stage="adult"
BASE COUNT 73 a 36 c 38 g 99 t
ORIGIN

Query Match 54.5% Score 21.8; DB 49; Length 246;
Best Local Similarity 78.8%; Pred. No. 3.9e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY
Db

4 TGTAAATCTATCTGCTATTCATGACAA 36
|||||
198 TGTCTACTTATCTGCTATTCATGACAA 230

RESULT 13

LOCUS T02774 280 bp DNA EST 29-NOV-1993
DEFINITION 011M7 gmbpFHB3.1, G. Roman Reddy Plasmodium falciparum genomic
clone PF011M, mRNA sequence.

ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM

T02774
T02774
9319290
T02774.1 GI:319290
EST.
malaria parasite P. falciparum.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 280)
Reddy, G.R., Chakrabarti, D., Schuster, S.M., Ferl, R.J., Almitra, E.C.
and Dame, J.B.

REFERENCE
AUTHORS
JOURNAL
MEDLINE
COMMENT

Gene sequence tags from Plasmodium falciparum genomic DNA fragments prepared by the genease activity of mung bean nuclease
Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)
94052193
Other ESTs: 011M3
Contact: Dame JB
Department of Pathobiology
University of Florida

2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880
Tel: 352 392 4700
Fax: 352 392 9704
Email: damel@cdr.flas.ufl.edu
Seq primer: T7.

FEATURES

source

Location/Qualifiers
1. 280
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone="PF011M"
/clone_1lb="gmbpFHB3.1, G. Roman Reddy"
/lab_host="E. coli XL1-Blue"
/note="Vector: pBluescript SK(+); Genomic DNA, from asynchronous blood stage parasites of the cloned Honduran HB3 isolate cultured in vitro, was digested with mung bean nuclease in the presence of 30% formamide at 50°C (Vernick, K.D., Imberiski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6895). The ends of the fragments were polished using T4 DNA polymerase, and the dephosphorylated pBluescript SK(+). Recombinant plasmids transformed E. coli XL1-Blue."

BASE COUNT
ORIGIN

104 a 23 c 30 g 109 t 14 others

Query Match 54.0% Score 21.6; DB 20; Length 280;
Best Local Similarity 75.0%; Pred. No. 4.3e+02;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 TGTAAATCTATCTGCTATTCATGACAAATG 39
|||||
Db 11 TTTAAATCTTTCCTTATTAAGAGAAAGATG 46

RESULT 14

LOCUS AA037681 316 bp mRNA EST 27-AUG-1996
DEFINITION ZK34911.1 Soares_pregnant_uterus_MDHPU Homo sapiens cDNA clone
IMAGE:484772 5', mRNA sequence.

ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM

AA037681
91512779
AA037681.1 GI:1512779
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 316)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
On Apr 14, 1993 this sequence version replaced gi:838110.

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LBNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from AmerSham
High quality sequence stop: 281.

FEATURES
source

1. 316
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:3758486"
/db_xref="taxon:9606"

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:19:58 ; Search time 329.34 Seconds
(without alignments)
366,951 Million cell updates/sec

Title: US-09-325-095-2425.SEQ

Sequence: 1 AAGATCAGACAAATGCTTCTTACCAATCA 38

Scoring table:

IDENTITY_MDC

Searched: 679419 seqs, 1590154680 residues

Database:

GenBank: 1: gb_ba1: 2: gb_ba2: 3: gb_om: 4: gb_ov: 5: gb_pat: 6: gb_ph: 7: gb_pl1: 8: gb_pl2: 9: gb_pl3: 10: gb_pl4: 11: gb_pl5: 12: gb_pl6: 13: gb_pl7: 14: gb_pl8: 15: gb_pl9: 16: gb_pl10: 17: gb_pl11: 18: gb_pl12: 19: gb_pl13: 20: gb_pl14: 21: gb_pl15: 22: gb_pl16: 23: gb_pl17: 24: gb_pl18: 25: gb_pl19: 26: gb_pl20: 27: gb_pl21: 28: gb_pl22: 29: gb_pl23: 30: gb_pl24: 31: gb_pl25: 32: gb_pl26: 33: gb_pl27: 34: gb_pl28: 35: gb_pl29: 36: gb_pl30: 37: gb_pl31: 38: gb_pl32: 39: gb_pl33: 40: gb_pl34: 41: gb_pl35: 42: gb_pl36:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.6	62.1	1251	1	SYODESA2 D13779 Synechococcus

2	22.8	60.0	204544	35	AC007677	AC007677 Homo sapi
3	22	57.9	144260	11	AC006299	AC006299 Homo sapi
4	22	57.9	258438	34	AC005046	AC005046 Homo sapi
5	21.8	57.4	2131	12	MMHCL	X66285 M.musculus
6	21.6	56.8	17696	7	SPBCL5C4	AL023290 S.pombe
7	21.6	56.8	116210	10	HS406C18	AL023290 S.pombe
8	21.6	56.8	17251	12	AB015623	AB015623 Mus muscu
9	21.6	56.8	149886	12	MMHCL38N12	AF049850 Mus muscu
10	21.4	56.3	212346	34	HS90G24	AF077409 Arabidops
11	21.2	55.8	68251	8	T24M8	AF024499 Caenorhab
12	21.2	55.8	37380	36	CEUFA2G2	Y09612 B.poriidilum
13	21	55.3	2460	1	PLPWRACG	AF053983 Podospora
14	21	55.3	2626	8	AF053983	AF053983 Podospora
15	21	55.3	90731	11	AC004897	AC004897 Homo sapi
16	21	55.3	5437	37	AF110008	AF110008 Molgula o
17	21	55.3	137611	42	AC006355	AC006355 Homo sapi
18	20.8	54.7	107848	8	ATAC007019	AC007019 Arabidops
19	20.8	54.7	183386	11	AC005229	AC005229 Homo sapi
20	20.8	54.7	176725	35	AC007386	AC007386 Homo sapi
21	20.8	54.7	80389	37	AC005643	AC005643 Drosophila
22	20.6	54.2	119707	7	PINCPTRPG	D17510 Pinus thunb
23	20.6	54.2	161516	35	AC007131	AC007131 Homo sapi
24	20.6	54.2	108320	35	AC007333	AC007333 Homo sapi
25	20.6	54.2	207188	35	AC007389	AC007389 Homo sapi
26	20.6	54.2	41396	36	DMC34F3	AL031583 Drosophila
27	20.4	53.7	14039	1	MTU19364	U19364 Methanobact
28	20.4	53.7	81729	7	AB016874	AB016874 Arabidops
29	20.4	53.7	37672	7	SCVCSMGN	X89633 S.cerevisia
30	20.4	53.7	6319	7	SCYOR291W	275199 S.cerevisia
31	20.4	53.7	6918	7	YSCRC1	D90459 Yeast R1C1
32	20.4	53.7	153472	9	AB020868	AB020868 Homo sapi
33	20.4	53.7	118968	10	HS796F18	AL031291 Human DNA
34	20.4	53.7	123397	11	AC004808	AC004808 Homo sapi
35	20.4	53.7	973	11	HSRPS14	AF058410 Homo sapi
36	20.4	53.7	339	14	G07818	G07818 human STS C
37	20.4	53.7	169	14	HMSX3123	L77630 Human chrom
38	20.4	53.7	3417	17	ARLVGPN	X52400 Nigeria Las
39	20.4	53.7	1830	17	LSVNG	K03362 Lassa virus
40	20.4	53.7	144952	34	HS417M14	AL024498 Homo sapi
41	20.4	53.7	193126	42	AC006045	AL006045 Homo sapi
42	20.2	53.2	44185	10	HS069729	U69729 Human Xp22
43	20.2	53.2	180664	34	LC406706	AC006706 Caenorhab
44	20.2	53.2	83527	37	L49408	L49408 Drosophila
45	20	52.6	3452	4	AF0101076	AF0101076 Gallus ga

ALIGNMENTS

RESULT 1
SYODESA2 1251 bp DNA BCT 03-FEB-1999
LOCUS Synechococcus sp. desa gene encoding delta 12 desaturase, complete cds.
ACCESSION D13779
NID 9488510
VERSION D13779.1 GI:488510
KEYWORDS delta 12 desaturase.
SOURCE Synechococcus sp. (strain:PCG7002) DNA, clone lib:lambd DASH II.
ORGANISM Synechococcus sp.
Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
REFERENCE 1 (bases 1 to 1251)
AUTHORS Sakamoto, T.
TITLE Direct Submision
JOURNAL Submitted (27-NOV-1992) to the DDBJ/EMBL/GenBank databases, Toshio Sakamoto, Graduate University of Advanced Studies, Molecular Biomechanics, Myodaiji, Okazaki, Aichi 444, Japan (Tel:0564-55-7602, Fax:0564-53-7400)
REFERENCE 2 (bases 1 to 1251)
AUTHORS Sakamoto, T.
JOURNAL Unpublished (1994)
REFERENCE 3 (sites)
AUTHORS Sakamoto, T., Wada, H., Nishida, I., Ohmori, M. and Murata, N.

```

* * * * *
* * 1 2104: contig of 2104 bp in length
* * 2105 2122: gap of unknown length
* * 2123 4081: contig of 1559 bp in length
* * 4082 4093: gap of unknown length
* * 4100 7841: contig of 3742 bp in length
* * 7842 7859: gap of unknown length
* * 7860 15888: contig of 8029 bp in length
* * 15889 15905: gap of unknown length
* * 15907 31646: contig of 15740 bp in length
* * 31647 31664: gap of unknown length
* * 31665 97469: contig of 65805 bp in length
* * 97470 97487: gap of unknown length
* * 97488 204544: contig of 107057 bp in length.

FEATURES
    source
        1..204544
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="NH0086N01"

BASE COUNT      61995 a 39664 c 38343 g 64434 t      108 others
ORIGIN
Query Match      60.0%; Score 22.8; DB 35; Length 204544;
Best Local Similarity 79.4%; Pred. No. 11;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY      1 AAGATCAGACAAATGCTAGCGCTTCTTTAGCC 34
||||| ||||||| || ||| || ||||
Db 64154 AAGGATAGAACAAATGCTAATAGCTGATGAGCC 64187

```

	AC006299/c				
LOCUS	AC006299	144260 bp	DNA	PRI	06-JAN-1999
DEFINITION	Homo sapiens, clone hRPK.17_A_1,				complete sequence.
ACCESSION	AC006299				
NID	g4106997				
VERSION	AC006299.1				
KEYWORDS	GI:4106997				
SOURCE	HTC.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
TITLE	Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
JOURNAL	1 (bases 1 to 144260)				
AUTHORS	Birten,B., Jinton,L., Nusbaum,C. and Lander,E.				
REFERENCE	Homo sapiens, clone hRPK.17_A_1				
AUTHORS	Unpublished				
	2 (bases 1 to 144260)				
	Birten,B., Jinton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,				
	Caster,J., Baldwin,D., Barna,N., Beckery,R., Benn,I., Brown,A.,				
	Battle,A., Cerny,T.J., Colangelo,M., Collins,S., Collamore,A.,				
	Cooke,P., Dearlano,K., Depayre,E., Devon,K., Dewar,K.,				
	Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,				
	Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Headford,A.,				
	Harems,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karalas,A.,				
	Lelocsky,J., Macdonald,P., Margusis,N., McEwan,P., McCurk,A.,				
	McGerran,K., Meldrum,J., Molina,M., Morris,W., Morrow,J.,				
	Mychalecky,J.J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,				
	Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A.,				
	Severin,P., Stange-Thomann,N., Stojanovic,N., Stone,C.,				
	Subramanian,A.A., Tesfaye,S., Tortorella-Miller,I., Vassiliev,H.,				
	Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.-J. and Zody,M.				
TITLE	Direct Submission				
JOURNALS	submitted (06-JAN-1999) Whitehead Institute/MIT Center for Genome				
COMMENT	Research, 320 Charles Street, Cambridge, MA 02141 USA				
	All repeats were identified using RepeatMasker: Smit, A.F.A. &				
	Green, P. (1996-1997)				
	http://ftp.genome.washington.edu/RM/RepeatMasker.html.				
FEATURES	Location/Qualifiers				
SOURCE	1..144260				
	/organism='Homo sapiens'				
	/db_xref='taxon:9606'				
	/clone='hRPK.17_A_1'				

```
/clone.lib="RPCI-11 human BAC library"
repeat_region 457..499 /rpt_family="AT_rich"
repeat_region 837..861 /rpt_family="AT_rich"
repeat_region complement(862..1010) /rpt_family="AluYo"
repeat_region complement(1011..1060) /rpt_family="(TAA)n"
repeat_region complement(1061..1182) /rpt_family="AluYo"
repeat_region 1183..1210 /rpt_family="AT_rich"
repeat_region 4144..4273 /rpt_family="L1ME"
repeat_region 5137..5273 /rpt_family="FLAM_C"
repeat_region 5300..5493 /rpt_family="L1MD"
repeat_region 5494..5775 /rpt_family="AluX"
repeat_region 5776..5796 /rpt_family="(TAA)n"
repeat_region 5797..6324 /rpt_family="L1MD"
repeat_region 6325..6543 /rpt_family="MER20"
repeat_region 6544..6756 /rpt_family="L1MD"
repeat_region 6763..6816 /rpt_family="AT_rich"
repeat_region 6925..7073 /rpt_family="AluYo/FRAM"
repeat_region 7086..7101 /rpt_family="L1MD2"
repeat_region 7339..8052 /rpt_family="L1MD2"
repeat_region 8056..8086 /rpt_family="L1MD2"
repeat_region complement(8515..8753) /rpt_family="L1MA9"
repeat_region 8878..8943 /rpt_family="purine-rich"
repeat_region 8952..9239 /rpt_family="AluSc"
repeat_region 9718..9775 /rpt_family="AT_rich"
repeat_region complement(10580..10601) /rpt_family="(TAAA)n"
repeat_region 11225..11247 /rpt_family="AT_rich"
repeat_region 11509..11949 /rpt_family="MT1H"
repeat_region 11988..12011 /rpt_family="(CA)n"
repeat_region 12097..12548 /rpt_family="MER65A"
repeat_region 13119..13157 /rpt_family="AT_rich"
repeat_region 13216..13240 /rpt_family="AT_rich"
repeat_region 13936..14549 /rpt_family="L1M4"
repeat_region 14550..14620 /rpt_family="MT2D"
repeat_region 14778..14893 /rpt_family="L1ME1"
repeat_region 15007..15035 /rpt_family="AT_rich"
repeat_region complement(15856..16793) /rpt_family="L1PA2"
repeat_region complement(19078..19390) /rpt_family="AluY"

repeat_region complement(19413..19720) /rpt_family="MSTO"
repeat_region 19768..19902 /rpt_family="MER91C"
repeat_region 19996..20035 /rpt_family="AT_rich"
repeat_region complement(20379..20919) /rpt_family="MT1D"
repeat_region complement(20988..20998) /rpt_family="MER5B"
repeat_region 20999..21145 /rpt_family="MER5A"
repeat_region complement(21146..21156) /rpt_family="MER5B"
repeat_region complement(21222..21657) /rpt_family="L2"
repeat_region complement(21729..22469) /rpt_family="L2"
repeat_region 22475..22520 /rpt_family="MADE1"
repeat_region 22569..22762 /rpt_family="MIR"
repeat_region 22953..23917 /rpt_family="L1MA7"
repeat_region complement(24118..24224) /rpt_family="MT1J"
repeat_region 24359..24404 /rpt_family="(CA)n"
repeat_region 24481..24516 /rpt_family="AT_rich"
repeat_region 24965..25007 /rpt_family="AT_rich"
repeat_region 26626..26830 /rpt_family="AT_rich"
repeat_region 27718..27944 /rpt_family="L1R33"
repeat_region 28087..28150 /rpt_family="MIR"
repeat_region complement(28175..28424) /rpt_family="MIR"
repeat_region 29543..29699 /rpt_family="L2"
repeat_region complement(30930..31229) /rpt_family="AluYa5"
repeat_region 31490..31846 /rpt_family="THE1B"
repeat_region 31992..32064 /rpt_family="MT1H"
repeat_region 32425..32532 /rpt_family="MIR"
repeat_region 33624..34049 /rpt_family="MSTB"
repeat_region complement(34212..34412) /rpt_family="MER20"
repeat_region complement(35457..35762) /rpt_family="AluSg"
repeat_region complement(36300..36505) /rpt_family="MT1J"
repeat_region 38065..38184 /rpt_family="L1PB1"
repeat_region complement(38455..38577) /rpt_family="L2"
repeat_region complement(39107..41490) /rpt_family="L1M4"
repeat_region 41504..41861 /rpt_family="MT1A1"
repeat_region complement(41874..42343) /rpt_family="L1M4"
repeat_region complement(42348..42710) /rpt_family="L1MD"
repeat_region 43747..44018 /rpt_family="AluX"
repeat_region 44098..44123
```

repeat_region /rpt_family="AT_rich"
complement(44124..44296)
repeat_region /rpt_family="Alusx"
complement(44297..44320)
repeat_region /rpt_family="(TAAA)n"
complement(44321..44448)
repeat_region /rpt_family="Alusx"
44701..44743
repeat_region /rpt_family="AT_rich"
complement(44395..45511)
repeat_region /rpt_family="L2"
complement(46038..46347)
repeat_region /rpt_family="Alusg"
46498..46606

Query Match 57.9%; Score 22; DB 11; Length 144260;
Best Local Similarity 73.7%; Pred. No. 24;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 AAGATCAGACATGCTTGGCTTTTTCATCATCA 38
111 111 111 111 111 111 111 111 111 111
DB 95158 AAGATCTGATCAATCTAGCTGTCTATCATCA 95121

RESULT 4
NC005046/c AC005046 238438 bp DNA HTG 12-JUN-1998
LOCUS Homo sapiens clone RG013F03, WORKING DRAFT SEQUENCE, 6 unordered
DEFINITION pieces.
ACCESSION AC005046
NID G3212945
VERSION AC005046.1 GI:3212945
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 238438)
REFERENCE 1 Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 238438)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 3296: contig of 3296 bp in length
3297 3313: gap of unknown length
3314 6202: contig of 2889 bp in length
6203 6219: gap of unknown length
6220 15488: contig of 9269 bp in length
15489 15505: gap of unknown length
15506 42224: contig of 26719 bp in length
42225 42241: gap of unknown length
42242 118044: contig of 75803 bp in length
118045 118061: gap of unknown length
118062 238438: contig of 120377 bp in length.
FEATURES
Source location/Qualifiers
1..238438
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RG013F03"
BASE COUNT 71781 a 48465 c 48870 g 69237 t 85 others
ORIGIN

Query Match 57.9%; Score 22; DB 34; Length 238438;
Best Local Similarity 83.3%; Pred. No. 24;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 AGAACATGCTTGGCTTTTTCATCATC 37
111 111 111 111 111 111 111 111 111 111
DB 19036 AGAACATGCTTGGCTTCTTATGCTTC 19007

RESULT 5
NMHC1/c NMHC1 2131 bp DNA ROD 28-OCT-1992
LOCUS M.musculus DNA for HC1 locus.
DEFINITION X66285
ACCESSION X61207
NID X66285.1 GI:51207
VERSION HC1 gene.
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2131)
REFERENCE 1 Kahn,T.
AUTHORS Kahn,T.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1992) T. Kahn, German Cancer Res Center, Im
Neuenheimer Feld 242, W-6900 Heidelberg, FRG
2 (bases 1 to 2131)
AUTHORS Kahn,T., Friesl,H., Copeland,N.G., Gilbert,D.J., Jenkins,N.A.,
Gissmann,L., Kramer,J. and zur Hausen,H.
TITLE Molecular cloning, analysis, and chromosomal localization of a
mouse genomic sequence related to the human papillomavirus type 18
E5 region
JOURNAL Mol. Carcinog. 6 (2), 88-99 (1992)
MEDLINE 93000417
COMMENT Similarity to the human Papillomavirus type 18 E5 region at the DNA
and putative protein level. Similarity of the putative protein
with the ORF of Q300 (Wagner, et al., J.Virol. 65:3529-3267(1991).
RNA overexpressed in mouse skin tumours obtained after carcinogen
(DMBA) or UV treatment. Amplification of the genomic DNA in the Eb
cell line and one UV induced mouse skin carcinoma.
FEATURES
Source location/Qualifiers
1..2131
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="8"
/cell_line="Eb"
/map="Jumb, Jund, linked to Plat, Flg"
1383..1739
CDS
/note="HC1 ORF; Protein sequence is in conflict with the
conceptual translation"
/codon_start=1
/protein_id="CAA46991.1"
/db_xref="PIR:S49330"
/db_xref="PIR:G1333929"
/db_xref="GI:1333929"
/db_xref="MGD:MGI:96220"
/db_xref="SPRTRMBL:002393"
/translation="RTYCIHKNSNRELLRASSRACHTSHITIHVDSFTSCVSY
CYCMCVYLDCSCMCVYACVCICVYVHVSVYACVCICVCMKVCACVCVPCVCVC
HMPNCHVCELCV"
misc-feature
1472..1691
/note="E5 similarity"
BASE COUNT 591 a 405 c 447 g 688 t
ORIGIN

Query Match 57.4%; Score 21.8; DB 12; Length 2131;
Best Local Similarity 78.8%; Pred. No. 36;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGATCAGACATGCTTGGCTTTTTCATCATC 33

DB 447 AATAACCAAAAGTCTAGGTTCTTTTC 415

RESULT 6

SPBC15C4 17696 bp DNA PLN 27-OCT-1998

LOCUS S. pombe chromosome II cosmid c15C4.

DEFINITION AL023290

ACCESSION 93116143

MID 93116143

VERSION AL023290.1 GI:3116143

KEYWORDS AGC1 homologue; amino acid permease; ATP dependent rna helicase; gcrany1geranyl transferase; zinc finger.

SOURCE fission yeast.

ORGANISM Schizosaccharomyces pombe

Eukaryota; Fungi; Ascomycota; Archizoscomycetes;

Schizosaccharomycetales; Schizosaccharomycetaceae;

Schizosaccharomyces.

1 (bases 1 to 17696)

Lyne, M., Rajandream, M.A., Barrell, B.G., Xiang, Z. and Aves, S.

Direct Submission

Submitted (05-MAY-1998) European Schizosaccharomyces genome

sequencing project, Sanger Centre, The Wellcome Trust Genome

Campus, Hinxton, Cambridge CB10 1SA, E-mail: Barrell@sanger.ac.uk

and Department of Biological Sciences, University of Exeter, Perry

Road, Exeter EX4 4QG, United Kingdom

Notes:

Details of yeast sequencing at the Sanger Centre are available on

the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/S_pombe/)

During 1995 to 1996 about 80% of S. pombe chromosome I was

sequenced by the Sanger Centre. The sequencing of the S. pombe

genome is now being continued with funding from The European

Commission. Fourteen European sequencing laboratories, including

the Sanger Centre, are participating in the project.

protein coding regions (CDS) have been predicted with the help of

computer analysis using the Genefinder program in Pombase (an ACCEB

database) with additional predictions for the branch-acceptor sites

supplied by the program Sp3splice. CAUTION: It is possible that for

any individual CDS we may have underestimated or overestimated the

number of introns/exons or we may not have chosen the correct

splice donor/acceptor sites. CDS are numbered using the following

system (eg SPBC25H2.01C: SP (S. pombe), B (chromosome 2), C25H2

(cosmid name), .01 (first CDS), C (complementary strand).

The more significant matches with motifs in the PROSITE database

are also included but some of these may be fortuitous. The length

in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the

sequenced clone. It may be shorter because we only sequence

overlapping sections once, or longer, because we arrange for a

small overlap between neighbouring submissions. Cosmid c15C4 is

overlapped at the 5' end by cosmid 14C8 and at the 3' end by cosmid

21H7.

FEATURES

Source

Location/Qualifiers

1..17696

/organism="Schizosaccharomyces pombe"

/strain="972h-"

/db_xref="taxon:4896"

/chromosome="II"

/clone="cosmid c15C4"

/map="f11R"

1..1901

/note="nominal overlap with SPBC14C8 S. pombe chromosome

2"

gene

complement(815..1928)

/gene="SPBC15C4.01C"

complement(join(815..1004,1082..1280,1332..1477,

1541..1684,1758..1928))

/gene="SPBC15C4.01C"

/note="SPBC15C4.01C, len:282, SIMILARITY: Homo sapiens,

Y103_HUMAN, hypothetical protein KIA0103, (297 aa),

fasta scores: opt: 350, E(): 2.9e-25, (33.3% identity, in

210 aa)"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAA1892.1"

/db_xref="PID:e1291650"

/db_xref="PID:g3116144"

/db_xref="GI:3116144"

/db_xref="SPREMBL:060110"

/translation="MSNSILKVPDQNPQIEVALFSQGEAYAKLGKXDEIMDYQKYF

IAALTGETVLAKKCKGNLRNDRFKSPRVEGLYGMFEATASEKDAASYNSKLSDEP

THVTYRKRLALRSWGQTKCEGLGILNYIDTFYNDLEAWEALADIVSYEAFESALF

CYEMRYLQPREPRFLRDGLYVLAQSNATNYSKAYCSVEICEYFQHMFTI

SKCCQQLERLKRLLKRLSKVNRKISDTPDESVALIQLSLKSDLPRAQKPKL

ALLKQC"

complement(1005..1019)

/gene="SPBC15C4.01C"

/note="ttacatttaattag, splice branch and acceptor"

complement(1076..1081)

/gene="SPBC15C4.01C"

/note="gtattt, splice donor sequence"

complement(1281..1296)

/gene="SPBC15C4.01C"

/note="ctattatgtagtag, splice branch and acceptor"

complement(1326..1331)

/gene="SPBC15C4.01C"

/note="gtaaagt, splice donor sequence"

complement(1478..1490)

/gene="SPBC15C4.01C"

/note="ctacttcacag, splice branch and acceptor"

complement(1535..1540)

/gene="SPBC15C4.01C"

/note="gtaaag, splice donor sequence"

complement(1685..1701)

/gene="SPBC15C4.01C"

/note="ttaatitcttaattag, splice branch and acceptor"

complement(1752..1757)

/gene="SPBC15C4.01C"

/note="gtaaagt, splice donor sequence"

2961..4994

/gene="SPBC15C4.02"

join(2961..3185,3230..3538,3652..4634,4727..4994)

/gene="SPBC15C4.02, len:594, SIMILARITY: Saccharomyces

cerevisiae, YLR253W, YL53_YEAST, hypothetical 65.9 kd

protein, (569 aa), fasta scores: opt: 1440, E(): 0, (44.5%

identity in 499 aa)"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAA1893.1"

/db_xref="PID:e1291651"

/db_xref="PID:g3116145"

/db_xref="GI:3116145"

/db_xref="SPREMBL:060111"

/translation="MSNYVLTSQVLYVYNSKRKYVWFMLRSFMSNRKFLVHLPL

HLNSLISICFLGEPGQELVASDHQFRNHRKPAVVAATTLMAVALVDPVKNAG

VSKRAYVYLAAGFLCPSDYKRYGSSVASEEENALISDCHLACRSKLGDEBENG

IYKIQGLSAMGVIPKREWTNTVVKLODRCPSYSLNDIDHLFVDRSGKLGDETFEP

DPILGVAASLAQVAKRLKSDVVAAYVQSPVSLNSPIDLSMTIRWVFAKIKTFEPD

EKLMLADEIERISLPDLDTREKNSITKREHRAHLSISLYPEVWMSKPRITIMYF

VAGARINDLSFMDHRSISROLVSYDCHITREMLFGGGLHCDPHGSGNYLRKPKN

KSPSPRNYEIVLDHGLRDIPHELVQVYANRWLNTINERKNLKFAKVAAYSDENP

PIFAVATIGRKYKLSKLPISGPNHBEOKRFISALQGGKLOSIIHLSTMPRLTLI

LMKNTDLVRSIDENLKTSGPEKLYLLINARCLRCVYDDKMSLMSRSFVSKVFG

TYRSLYFSSLSKXTYLAENFEYVYHML"

3186..3191

/gene="SPBC15C4.02"

/note="gtattt, splice donor sequence"

3219..3229

/gene="SPBC15C4.02"

/note="ctattttag, splice branch and acceptor"

3539..3544

/gene="SPBC15C4.02"

/note="gtaaagt, splice donor sequence"


```

repeat_region 12509..12586
/note="MER4A repeat: matches 228..308 of consensus"
12587..12890
/note="MER4A repeat: matches 367..661 of consensus"
repeat_region complement(12882..13148)
/note="MIR1D repeat: matches 287..1 of consensus"
repeat_region complement(13162..13209)
/note="MIR repeat: matches 138..90 of consensus"
13437..15600
/note="LIP4 repeat: matches 730..893 of consensus"
repeat_region complement(16025..16313)
/note="AluX repeat: matches 294..3 of consensus"
repeat_region 19307..19348
/note="21 copies 2 mer aa 81% conserved"
19388..19775
/note="MIR1C repeat: matches 58..456 of consensus"
20264..20789
/note="LIM4 repeat: matches 514..1045 of consensus"
repeat_region complement(20866..21101)
/note="LIM1 repeat: matches 1051..614 of consensus"
21096..21220
/note="LIM2 repeat: matches 667..796 of consensus"
repeat_region complement(21209..21362)
/note="MER45 repeat: matches 168..11 of consensus"
repeat_region complement(23167..23470)
/note="MIR1D repeat: matches 455..154 of consensus"
23522..23571
/note="25 copies 2 mer tt 82% conserved"
24349..24442
/note="MIR repeat: matches 80..178 of consensus"
repeat_region complement(24458..24683)
/note="MIR repeat: matches 256..21 of consensus"
misc_feature complement(25195..25446)
/note="match: GGG B34150"
25885..25914
/note="15 copies 2 mer tt 87% conserved"
26931..27144
/note="LIP4 repeat: matches 678..891 of consensus"
27916..28087
/note="MER20 repeat: matches 1..178 of consensus"
28842..28889
/note="24 copies 2 mer gt 100% conserved"
29010..29490
/note="MIR1D repeat: matches 1..505 of consensus"
29816..29924
/note="MIR repeat: matches 1..114 of consensus"
repeat_region complement(31395..31696)
/note="AluX repeat: matches 302..1 of consensus"
32451..33106
/note="LIP9 repeat: matches 238..905 of consensus"
33416..33594
/note="MIR repeat: matches 70..252 of consensus"
33740..33927
/note="LIM7 repeat: matches 846..1034 of consensus"
33997..34095
/note="MER5A repeat: matches 4..109 of consensus"
repeat_region complement(34339..34686)
/note="MIR1A1 repeat: matches 365..1 of consensus"
repeat_region complement(35068..35168)
/note="MER5A repeat: matches 189..83 of consensus"
/note="MIR repeat: matches 151..114 of consensus"
35515..35672
/note="MIR repeat: matches 54..206 of consensus"
36412..36759
/note="LIP5 repeat: matches 545..892 of consensus"
36983..37010
/note="14 copies 2 mer aa 89% conserved"
repeat_region complement(37417..37465)
/note="MIR2 repeat: matches 146..95 of consensus"
repeat_region complement(37765..37929)
/note="MER45 repeat: matches 178..15 of consensus"
misc_feature complement(38232..38332)

```

```

repeat_region /note="match: SRS L42655"
38603..38724
/note="MIR2 repeat: matches 22..146 of consensus"
repeat_region complement(38763..39144)
/note="LIMC2 repeat: matches 1075..665 of consensus"
repeat_region complement(39256..39911)
/note="LIM6 repeat: matches 1047..312 of consensus"
repeat_region complement(40016..40315)
/note="LIMC1 repeat: matches 669..370 of consensus"
40587..41141
/note="LIP14 repeat: matches 338..896 of consensus"
44248..44562
/note="AluX repeat: matches 1..302 of consensus"
repeat_region complement(44565..44715)
/note="AluO repeat: matches 281..122 of consensus"
repeat_region complement(45854..45954)
/note="LIM3 repeat: matches 338..238 of consensus"
45959..45980
/note="11 copies 2 mer at 100% conserved"
repeat_region complement(46036..46247)
/note="LIM5 repeat: matches 218..7 of consensus"
46271..47164
/note="MER21A repeat: matches 25..902 of consensus"
repeat_region complement(47170..49364)
/note="L1 repeat: matches 3872..1681 of consensus"
49365..50875
/note="L1 repeat: matches 3874..5390 of consensus"
50733..51615
/note="LIP2 repeat: matches 1..890 of consensus"
51852..51885
complement(52467..53595)
/note="L1 repeat: matches 5223..4076 of consensus"
53604..54173
/note="LIM10 repeat: matches 334..912 of consensus"
54174..54207
/note="17 copies 2 mer ca 100% conserved"
54240..54388
/note="LIM9 repeat: matches 902..1049 of consensus"
repeat_region complement(54850..55704)
/note="MIR2-internal repeat: matches 5214..4354 of consensus"
repeat_region complement(56831..56959)
/note="MIR2-internal repeat: matches 2337..2406 of consensus"
57082..57127
/note="23 copies 2 mer gt 100% conserved"
repeat_region complement(57128..57288)
/note="LIM2 repeat: matches 896..760 of consensus"
repeat_region complement(57268..57589)
/note="AluX repeat: matches 302..1 of consensus"
repeat_region complement(57616..58103)
/note="LIM2 repeat: matches 724..227 of consensus"
complement(58189..58402)

```

```

Query Match 56.8% Score 21.6; DB 10; Length 116210;
Best Local Similarity 75.0%; Pred. NO. 36;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

OY 1 AAGGATGAGACATGCGTTCCTTACCAT 36
DB 60604 AATGCCAGACAGTACCTTGGCTTTTACCAT 60569

```

```

RESULT 8
LOCUS AB015623 17551 bp DNA ROD 30-JUN-1998
DEFINITION Mus musculus Cyp21 and Slp genes for steroid 21-hydroxylase and
AB015623 sex-limited protein, complete and partial cds.
ACCESSION 63273213
NID 63273213
VERSION AB015623.1 GI:3273213
KEYWORDS X: Tnx; tenascin-X; Slp; sex-limited protein; steroid
21-hydroxylase.

```

SOURCE Mus musculus (strain: F1 animals of C57BL/6 and CBA cross)
cell_line: T12 DNA.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 17251)
AUTHORS Matsumoto, K. and Ikuta, T.
JOURNAL Direct Submission
Submitted (15-JUN-1998) to the DDBJ/EMBL/GenBank databases.
Ken-ichi Matsumoto, Graduate School of Pharmaceutical Sciences,
Hokkaido University, Department of Molecular Biology; Kita 12,
Nishi 6, Kita-ku, Sapporo, Hokkaido 060-0812, Japan
(E-mail: kmatsumapharm.hokudai.ac.jp, Tel: 81-11-706-3731,
Fax: 81-11-706-4988)
2 (bases 1 to 17251)
AUTHORS Matsumoto, K. and Ikuta, T.
JOURNAL Mus musculus 5' truncated pseudogene of tenascin-X, steroid
21-hydroxylase (Cyp21), and sex-limited protein (slp) genes,
partial cds and complete sequences
Published Only in Database (1998) In press
Interpreted repeats and simple sequence repeats were identified
with RepeatMasker accessed through
http://ftp.genome.washington.edu/cgi-bin/RepeatMasker.html.
COMMENT location/Qualifiers
1. 17251
/organism="Mus musculus"
/strain="F1 animals of C57BL/6 and CBA cross"
/db_xref="taxon:10090"
/cell_line="T12"
/chromosome="17"
/note="1. 3793: genomic PCR product; our used primer set
is as follows: forward primer corresponds to nt 58711 to
58735, and reverse primer to nt 63202 to 63222. Each nt
number is based on the sequence from Accession
No. AB010266.; 1632. 17251: A mouse genomic phage clone:
this clone was obtained by the screening with a mouse TNX
cDNA probe, 800-bp Hind III fragment which encodes FNIII
repeat M26-M28."

misc_feature 260..538
/note="high similarity to nt 58991-59269 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 1229..1511
/note="high similarity to nt 60069-60350 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 2322..2630
/note="high similarity to nt 61162-61470 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 2935..3265
/note="high similarity to nt 61162-61470 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 3516..3638
/note="high similarity to nt 61780-62115 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 3737..3877
/note="high similarity to nt 63265-62487 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 3967..4097
/note="high similarity to nt 63378-63508 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 4220..4352
/note="high similarity to nt 63631-63763 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 4446..4597
/note="high similarity to nt 63862-64013 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 4698..4794
/note="high similarity to nt 64114-64210 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 4882..5043
/note="high similarity to nt 64297-64458 of mouse
tenascin-X (GenBank Accession Number AB010266)."
misc_feature 5125..5288
/note="high similarity to nt 64540-64703 of mouse
tenascin-X (GenBank Accession Number AB010266)."

exon complement(5441..6120)
/gene="Cyp21"
/number=10
complement(5441..8490)
/gene="Cyp21"
complement(5441..5446)
/gene="Cyp21"
/note="This polyA signal is used for Cyp21 gene."
5590..5845
/gene="Cyp21"
/note="high similarity to nt 65005-65293 of mouse
tenascin-X (GenBank Accession Number AB010266)."
complement(join(5852..6120,6205..6308,6378..6547,
6744..6944,7058..7144,7233..7334,7421..7516,7622..7764,
8125..8214,8279..8480))
/gene="Cyp21"
/note="similar to mouse steroid 21-hydroxylase A (GenBank
Accession Number M15009) or mouse steroid 21-hydroxylase B
(GenBank Accession Number M15008)."
/codon_start=1
/product="steroid 21-hydroxylase"
/protein_id="BAA31153.1"
/db_xref="PID:d1032100"
/db_xref="PID:g33273214"
/db_xref="GI:3273214"
/translation="MLPGLLLLLLAGTRWLGQWLRKRLHPPLAPGLFLQPN
LPVILGTLQGLPIYRIILGMDVVLNSKRTIEALLQKWDVFAQPHNLKMLD
LSLGDYSIMKAKRLSRSLMGLMDSKPELLEQLQECERRAQAGPVAKHE
FSLNGSISLCTFGDKDSTVORLHDCVODLLQAWHMSIQILITLPLRFLNPF
OKTQIOESRPHIYKQDLRKRDLSVAGQDIDVYLQCYEKORQDGRERLHGIV
HMSYVDPLFGTETRTATLLSNAYVFLHHPETQRLDEDLKLGPSQILRRMKL
PLNATTAIEVRLRPVPLALPHATRSASSISGYDIDKDWIILPNIGCANLDEWVWL
PSKWPREFLEPGNRPPTSPSCGRCARVCLGEPRLRLFPVLLARLQLAFTLLPPDGT
LPSQDPYAGINLPIPFQYRLQPRMLAPDQGERP"
complement(6205..6308)
/gene="Cyp21"
/number=9
complement(6378..6547)
/gene="Cyp21"
/number=8
complement(6744..6944)
/gene="Cyp21"
/number=7
complement(7058..7144)
/gene="Cyp21"
/number=6
complement(7233..7334)
/gene="Cyp21"
/number=5
complement(7421..7516)
/gene="Cyp21"
/number=4
complement(7622..7764)
/gene="Cyp21"
/number=3
7829..7922
/rpt_family="B1-F"
7916..8010
/note="purine-rich"
8068..8097
/note="POLK_G"
complement(8125..8214)
/gene="Cyp21"
/number=2
complement(8279..8490)
/gene="Cyp21"
/number=1
8438..8479
/rpt_unit="CAG
complement(8963..9096)
/rpt_family="PB1"
complement(9170..9346)


```

repeat_region      /rpt_family="B4"
                    9348..9538
repeat_region      /rpt_family="B3"
                    /note="HERV"
                    complement(9580..9922)
repeat_region      /note="POLY_A"
                    complement(10266..10296)
repeat_region      /rpt_family="B3"
                    complement(10409..10445)
repeat_region      /rpt_family="B3"
                    complement(10499..10835)
                    /note="MMVL30"
repeat_region      complement(10839..11036)
repeat_region      /rpt_family="B3"
                    complement(11062..11191)
                    /rpt_family="PB1D7"
repeat_region      11226..11295
                    /rpt_family="B4A"
                    11242..11346
repeat_region      /rpt_family="PB1D9"
                    11352..11415
repeat_region      /rpt_family="ID5"
                    complement(11583..11649)
repeat_region      /rpt_family="MIR"
                    11685..11735
                    /note="A-rich"
repeat_region      12189..12281
                    /rpt_family="PB1D10"
                    12315..12327
repeat_region      /rpt_family="B4"
                    12328..12385
repeat_region      /rpt_family="ID3"
                    12386..12423
                    /rpt_family="B4"
                    complement(13200..13440)
                    /note="high similarity to nt 18060-18300 of mouse Slp gene
                    (GenBank Accession Number M64933)"
exon                complement(13216..13221)
                    /note="This polyA signal is used for SLP gene."
polyA_signal        complement(13299..16983)
gene                /gene="SLP"
                    complement(13299..13440,13577..13709,13994..14077,
                    14212..14310,14388..14477,14599..14701,15215..15289,
                    15419..15509,15604..15784,15871..15964,16924..16983))
                    /gene="SLP"
                    /codon_start=1
                    /product="sex-limited protein"

Query Match      56.8%; Score 21.6; DB 12; Length 17251;
Best Local Similarity 75.0%; Pred. No. 39;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

07 1 AAGGATCAGACATGCTAGCTTCTTTCATCCAT 36
 DB 976 AAGTACGACAAATTTCCATAGCTTCTTTCATCCAT 941

RESULT 9
 LOCUS MMHC438N12/c DNA ROD 10-MAR-1998
 DEFINITION Mus musculus major histocompatibility locus class III region:
 complement C4 (C4) and cytochrome P450 hydroxylase A (CYP210H-A)
 genes, complete cds; slp pseudogene, complete sequence; NG6, SKI,
 and complement factor B (Bf) genes, complete cds; and complement
 factor C2 (C2) gene, partial cds.
 ACCESSION AF049850
 NID AF049850.1 GI:2944418
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1. (bases 1 to 149886)

```

AUTHORS Rowen,L., Qin,S., Lasky,S.R., Loretz,C., Dors,M., Mahairas,G. and
          Hood,L.
TITLE Sequence of the mouse major histocompatibility locus class III
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149886)
AUTHORS Rowen,L.
TITLE Direct Submission
COMMENT Submitted (23-FEB-1998) Department of Molecular Biotechnology, Box
          357730 University of Washington, Seattle, Washington 98195, USA
          This sequence overlaps BAC 29N7, found in GenBank Accession Number
          AF030001, by 93 bases on the 3' side of 29N7. Sequencing
          methodology: high redundancy shotgun. Interspersed Repeats were
          identified with RepeatMasker (available from
          http://ftp.genome.washington.edu/RM/RepeatMasker.html) Simple
          sequence repeats were identified with spunik (available from
          http://www.geospiza.com/people/chris/software/spunik.html).

FEATURES
          Location/Qualifiers
            source          1..149886
                           /organism="Mus musculus"
                           /strain="129"
                           /db_xref="taxon:10090"
                           /chromosome="17"
                           /clone_lib="Genome Systems BAC Library"
                           /haplotype="H2-B"
                           1..93
                           /note="Overlap with BAC 29N7, GenBank Accession Number
                           AF030001, at positions 201873-201864"
                           1..20297
                           /note="this span duplicates the region between 80946 and
                           99794; it contains the C4 gene and a copy of the 3' end of
                           the G7 gene, exons 5-7"
                           1..3119
                           /organism="Mus musculus"
                           /db_xref="taxon:10090"
                           /clone="BAC 232L6"
                           /sub_clone="PCR product"
                           complement(27..313)
                           /rpt_family="MMVL30"
                           complement(316..423)
                           /rpt_family="(GGA)n"
                           370..450
                           /note="complex (CT)n microsatellite repeat -- low quality
                           data"
                           complement(414..543)
                           /rpt_family="PB1D7"
                           596..700
                           /rpt_family="PB1D9"
                           706..763
                           /rpt_family="ID4"
                           939..971
                           /rpt_family="POLY_A"
                           1941..1998
                           /rpt_family="PB1D10"
                           2899..149886
                           /organism="Mus musculus"
                           /db_xref="taxon:10090"
                           /clone="BAC 438N12"
                           complement(2992..18332)
                           /gene="C4"
                           /note="complement C4 gene"
                           complement(join(2992..3133,3271..3403,3688..3771,
                           3906..4004,4082..4171,4310..4412,4926..5000,5130..5220,
                           5315..5495,5582..5675,6676..6735,7230..7397,7479..7711,
                           7906..8077,8166..8282,8378..8534,8638..8713,8872..9081,
                           9207..9296,9383..9434,9910..10113,10242..10381,
                           10571..10641,10746..10860,10953..11147,11401..11475,
                           12139..12265,13349..13507,14161..14346,14442..14627,
                           14773..14949,15035..15150,15468..15600,15676..15781,
                           16297..16393,16494..16576,16870..16958,17038..17108,
                           17317..17518,17975..18167,18268..18332))
                           /gene="C4"
                           /note="intron-exon boundaries defined in relation to cDNAs

```


contaminated with foreign sequence from E. coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: dj90G24 Contig_ID: 05038 acc=AL008723 Length: 2325 bp Unfinished: dj90G24 Contig_ID: 04825 acc=AL008723 Length: 4169 bp Unfinished: dj90G24 Contig_ID: 04564 acc=AL008723 Length: 1015 bp Unfinished: dj90G24 Contig_ID: 04865 acc=AL008723 Length: 4623 bp Unfinished: dj90G24 Contig_ID: 04788 acc=AL008723 Length: 3973 bp Unfinished: dj90G24 Contig_ID: 05022 acc=AL008723 Length: 1609 bp Unfinished: dj90G24 Contig_ID: 00436 acc=AL008723 Length: 2078 bp Unfinished: dj90G24 Contig_ID: 04713 acc=AL008723 Length: 1701 bp Unfinished: dj90G24 Contig_ID: 04934 acc=AL008723 Length: 9778 bp Unfinished: dj90G24 Contig_ID: 04936 acc=AL008723 Length: 1537 bp Unfinished: dj90G24 Contig_ID: 05387 acc=AL008723 Length: 2260 bp Unfinished: dj90G24 Contig_ID: 04325 acc=AL008723 Length: 168478 bp.

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will be preserved.

FEATURES

source

1. .212346
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="90G24"

BASE COUNT 55209 a 45984 c 45491 g 56862 t 8800 others

ORIGIN

Query Match 56.3%; Score 21.4; DB 34; Length 212346;
Best Local Similarity 80.6%; Pred. No. 42;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

7 CAGACATGCGCTTCTTACGATC 37
||||||| ||||| ||||| ||||| |||||
Db 105398 CAGACATGCTTCTTACGATC 105428

RESULT 11

LOCUS T24M8 68251 bp DNA 04-DEC-1998
DEFINITION Arabidopsis thaliana BAC T24M8.
ACCESSION AF077409
NID G3319365
VERSION AF077409.1 GI:3319365
KEYWORDS
SOURCE
ORGANISM

thale cress.
Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.

1 (bases 1 to 68251)
Washington University Genome Sequencing Center.
The A. thaliana Genome Sequencing Project
Unpublished (1997)

2 (bases 1 to 68251)
Latreille, P., Elliott, G. and Le, T.
The sequence of A. thaliana T24M8
Unpublished (1998)

3 (bases 1 to 68251)
Waterston, R.
Direct Submission
Submitted (10-JUL-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 68251)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted by: Genome Sequencing Center

Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: twilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by
M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is T24H24, 200 bp overlap; 3' cosmid is T25H8.
Actual start of this cosmid is at base position 72886 of T24H24;
actual end is at 68251 of T24M8.

NOTES:

Coding sequences below are predicted from computer analysis, using
the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES

source

1. .68251
Location/Qualifiers

/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/clone="T24M8"
/chromosome="IV"
/map="unknown"

complement(467..2435)
/gene="T24M8.2"
complement(join(467..713,832..1734,1843..2435))
/gene="T24M8.2"
/note="similar to maize transposon MDR mudra protein
(GB:AL021710)"

/codon_start=1
/evidence-not_experimental
/protein_id="AAC28217.1"
/db_xref="PDB:G3319368"
/db_xref="GI:3319368"

/translation="MFRVSLFQMPRLTSITIQACVETSSIFDQVSSAFSTGSD
ADSLFTKRRKDEMYFTLMPKRVKHSFEFFHVKSDLIRVLAHIDENCKMLRAK
AGSSESVIRKRVSHSODSLRNVSHRQFAPRLRLINHLGKGLPLGPOLEI
FRDHGAGIVSKAMRVOEHVELARGLPDSEFVLPGRKLRKYSIDGAMHLSKF
KGLTGASADGNENLPSAFATVDEFNDSMPFLKCLNIIPDEKLDVSDRAAS
IAGSLSENYPLAHGLCTFHLQKLEHFGSSLIIPVYASVRYTETEDSLFKIT
NSDKTLAQYLYEVDYRKMRSRAYSPSNRYNTNLSAEVALLKQNEYRINCLFESI
RSIMTRFNERREBSOHPSAVITINVKRKAISYDISTRLEVCQVQDEEFYKGDIK
THLVNLDKRTCTCMEDIDKFPYAHGJASKRHINLNKMFVDFEHSYRNRQAVSEI
HPNGDMYEIRIPDYSVEVLCIPSTRVPSGRKKRKRIPSVWEGRSQPKRLKCSRC
GSGGHNKSTCVAA1"

misc_feature

gene
complement(16052..20499)
/gene="T24M8.3"

complement(join(16052..16549,16634..16718,16801..16994,
17108..17161,17278..17361,17508..17588,17713..17831,
17882..17954,18017..18097,18226..18339,19111..19260,
19333..19446,19551..19794,19838..20049,20149..20499))
/gene="T24M8.3"

/note="contains similarity to human CD44R1 gene product"

```

(CB:X56794)"
/codon_start-1
/evidence-not_experimental
/protein_id="AAC28218.1"
/db_xref="PID:93319369"
/db_xref="GI:3319369"
/translation="MADIDKFLSPMGRESFLKAVGRPPKRLNGKSTGRQLTLD
PKSLVSQLQKTFRLKGPALQDLIAFQSWAEDVEYDKRVAAMDILVSHQKQ
VLSLPQIHALLSNNLKPVDPLIVGDLNMAQSWAEDVEYDKRVAAMDILVSHQKQ
KIDMGGDAITLPIISVYKRGKRGKSGSPRSTRSKPELINSDEKSLISLQKRE
EASNRMKLEKAKTKYRFRKSTKSSFYACSSISGRKKIMEVPIOSQTSSEHTPEH
VVGQPEVAGQVDDHLDKQDSSDSEPIINTEITDDPMQVFTVPIQSESNQDNDGK
NVPYDPTVQDQANEDVDSDQDASERPTTSGDLPKRENSSELETNAETVDC
KQDSDIRETASHSHIDLPIDQSSSQAKINSELIMFKQVNLMSSEINFAANNIT
LCLSHIFVYNIQDTMEAAASHLGFAPKAKRVLESAAVQALSLQDPEVNF
DQTKMAYIYSHPIDTNGKPKNKQLESPTPEADITDSODIGADLMDQDEGVVD
SDSPAREKERTLSEAEVLYAELLSEKTSYELLPMSSEFALFENITSAAPT
EHLTSSGYLISNKPILLSAKPTNMYSTLMEVYLSLQKLTATTLNORAAVQPMFA
NHLQKIKSFKAAMKSRKNSPKQPTVSGPTWEPADIDITYIPMTDSKRWYGLA
INLGVMSYEIDPNTDLIEDEVRKFIIEVYVIMPEYLIQRTCKPDCSONHGLOPYTK
RLDGLYKRLRSQDCGPVAM"
complement(23787..26319)
/gene="T24M8.4"
complement(join(23787..23958,25017..25198,25452..25499,
25812..25917,26126..26319))
/gene="T24M8.4"
/notes="contains similarity to C3HC4-type zinc fingers
(pfam: zf-C3HC4.hmm, score: 32.94)"
/codon_start-1
/evidence-not_experimental
/protein_id="AAC28219.1"
/db_xref="PID:93319370"
/db_xref="GI:3319370"
/translation="MLESYAGDVHTVPOSENSMEERGGEIVPPLNFAVNDGIFR
SGFEPVSFSLQSLRLKSIYICPEAYEVREFAKSGIOPFQGLERCYRVEP
FVNIDPEYRREALQRTGCLVGCVRKIQWCLSTIDEYQRAAKARISDRREHLE
DISNKLKPLTSPSCSRGVNNDSCVCKGRKKAFLPCGHTFCVCSREYWLNRGS
CPLNRPITIEITDIY"
complement(32861..34474)
/gene="T24M8.5"
complement(join(32861..33058,33121..33174,33274..33499,
33597..33641,33782..33824,34028..34171,34424..34474))
/gene="T24M8.5"
/notes="similar to chloroplast triose phosphate
translocators"
/codon_start-1
/evidence-not_experimental
/protein_id="AAC28223.1"
/db_xref="PID:93319374"
/db_xref="GI:3319374"
/translation="MISSIRPVLPBLNIVGIGIYFAIMWALNGVFNYYNKKVLANP
YLMVLTLSLACGSLMLVSWVALAHTIGHVAYIRCALAAMELNFMNMGMSMIS
NLAFFRNIFSKGKMGKGSVYNTYACLSMNSLITVPFASVAGCPOMWAGMNDV
SKSDOTLSKMWAVSHVYHLVNOVSYIPRCNLHLLPPLKRVNLAIAIILGFIY
SOIKRVRKNHLLVLCIGMLEPIYITL"
complement(38759..40368)
/gene="T24M8.7"
complement(join(38759..38977,39025..39081,39131..39242,
39720..39951,39999..40140,40198..40368))
/gene="T24M8.7"
/codon_start-1
/evidence-not_experimental
/protein_id="AAC28222.1"
/db_xref="PID:93319373"
/db_xref="GI:3319373"
/translation="MNFYKDGKRVISGKHEGLIYLEGVYVAAKESVPAENGKSVY
KAVGVRRKSKSESSVGGDSSEPSDCLVEVSESGPSHNDPLLRHKKMMF
CDKRGIVYEMKELEKASRIQGMENRKRKAKVYLLDLDESKETKTKMYATIVGT
NITSQIENLNRREGIGRNLPRRSRWCETKRRRQGRKTLVREGSLYSVVLGL
MINQTNPLKSLISLSEFTLKLGLSVAFAGSHSSNSLILSLKGLHLHVENRKR
WGKETMADGEFVCSPEVGVWCMCFHDMKR"
complement(41605..43570)
/gene="T24M8.6"

```

CDS

```

complement(join(41605..42323,42398..42665,42947..43570))
/gene="T24M8.6"
/notes="Similar to pectinesterase; T24M8.6"
/codon_start-1
/evidence-not_experimental
/protein_id="AAC28220.1"
/db_xref="PID:93319371"
/db_xref="GI:3319371"
/translation="MLVVESEFLLIMTVYIGVSKYECDDKHSCONFLELTKTASSS
LSEIRRELLIIVKNSVRIKIDMAMIGMEDTKOHEEMENDLCLKEDTNLFEDMES
AKDRIIRREVLIGEPYIGSYENIHWTLSGLSYTTCIDELGDCAYKRVREPOLO
DLISAKAALALFISISRDNTELVNSVPNSPULSHYDKDILNBOALAKIADIV
YAKDGTGKTVNNAIAPQHSKRRTITIKTIDELVATEWTKPMTLIGGQDS
TITIGNSASVNRRTFYATDNSNRTKMEEASNGGFIQVDCFRVTGVPAGPAV
ALRPMGDSNFTIYQSRKEDKSGFSIONCNITASSDLDATATYKTLGRWRIFSTY
AVLOSIEDLDPDPMGWEEETLSLTHREYQNRGAVTSRNVMSGPKVVKDKK
QATPEYVAKLIDGFTWLKESRIPKSGI"
complement(47434..56353)
/gene="T24M8.8"
complement(join(47434..47887,53313..56353))
/gene="T24M8.8"

```

gene

CDS

```

Query Match          55.8%  Score 21.2: DB 8: Length 68251;
Best Local Similarity 76.5%: Pred. No. 54;
Matches 26: Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY      5 ATCAGACAAATGCTAGCGCTTCCTTACCATCA 38
DB 18646 AACACAGAAATCAAGCCTTCTTACCATCA 18613
      1 ||||| | ||||| ||||| |||||

```

RESULT 12

```

CELFA2G2/C          CELFA2G2  37380 bp  DNA      INV      16-SEP-1997
LOCUS              Caenorhabditis elegans cosmid F42G2.
DEFINITION         AF024489
ACCESSION          AF024489
VERSION            AF024499.1 GI:2394467
KEYWORDS
SOURCE
ORGANISM           Caenorhabditis elegans strain-Bristol N2.
                  Caenorhabditis elegans

```

```

REFERENCE
AUTHORS            Bonfield,J., Burton,J., Connell,M., Cogsey,T., Cooper,J.,
                  Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
                  Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
                  Johnston,L., Jones,M., Kersey,J., Kirsten,J., Laister,N.,
                  Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
                  O'Callaghan,M., Parsons,J., Percy,C., Rikken,J., Roopra,A.,
                  Saunders,D., Showkhen,R., Smalton,D., Smith,A., Sonhammer,E.,
                  Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaundin,M.,
                  Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
                  Wilkinson-Sproat,J. and Wohlman,P.

```

TITLE

Nature 368 (6466), 32-38 (1994)

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted by: Gene Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: twenematode.wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is F30D4, 200 bp overlap. Actual start of this cosmid is at base position 1 of CELF42G2; actual end is at 37407 of CELF42G2.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES

SOURCE

1. 37380

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="II"

/clone="F42G2"

/complement(508..6271)

/gene="F42G2.4" (508..777,833..1484,1530..1648,1721..1884,2233..2446,2499..2768,3118..3543,5021..5200,5246..5864,6162..6271)

/gene="F42G2.4"

/codon_start=1

/evidence-not_experimental

/protein_id="AAB70356.1"

/db_xref="PID:g2394470"

/db_xref="GI:2394470"

/translation="MNVSELFSEMOIVAPLTLLNPLQVAVOVLKLTGVTDLASKVC

RGLSAYDKLPLANSIRFQWKNDEVRHNLGIEIKITPEKPVDTIYAVNGLEKALP

GENETRIAFKDLBGLKHAITELSLYNNRDLIENSFEEALRTNKPVHIKSLYDES

INGIFKLLQYDDVLEDELEFEVSNNGFEKINSIDQWNAKSTFEGSCAFDIKEIE

HLFHEKILLITANKFTVDAVKIRIDDLKSTFECSTYFEGSNPKIKAVFQDPF

TEGNLMSHGSKNFKLHFAVNFSCIKIMSYAVNLCOATNEMILTKLVDKISGVY

SOIRDKGELLIFEDKSAFVYKYEYDAGDNTVYREVKLVQEKDPAVDCEDNHI

LNNATLADISFSLIADNTNDNVEADRVHQRKSLQKQKGPVATTLKLDNIS

KLRVQLQDFDTLTLELINDTNKHKKEITFEDIVEMQWNAKKEINLNFVSI

FEHFSCHESAVVEWIKISADIIKLIEDFNAPEIRRIIVKYSIDVTKLKLFEIG

TYDMNRHVSINNNEEFQCTLANOFYIAHPYGFSDQNSRNFAPFLSYKLHID

SCRSCTROKFSATRLVAGRTTLMETLLKVKLCONEAIAAGRLSLSTSGPTM

LHCIVMGKTOYIVLKHLYLRVRSRAIROCIDIKDPVAVATTIYGGTSTVEI

FGKDSKLFGISGSKSDQTNKLDISNFSNMLLIKOKKALTELEIYDEERKES

EHLIIECEEFVLPDLRLEVEKFTVNRNPEIIGQLTYMDPVSLRIEISEVNR

PMKQIEIDNKINTLNMENHAKKELTINNTVKKDVEVLHIFSEVDILLDMLONV

YLKDTPLFSPTFCOKAFQVNSDINETIFVDTFLDPSRVSDSSGLPNIWFESYAD

VDELDTIYVEROEFMTRIPIRETRRQHFHFRL"

/complement(7207..8767)

/gene="F42G2.5" (7207..7250,7300..7600,7776..7830,

7885..8003,8052..8197,8245..8535,8599..8767)

/gene="F42G2.5"

/note="similar to Aplysia californica vesicle-associated

membrane protein/synaptobrevin binding protein

(GB:U03679); also contains similarity to C3HC4-type zinc

misc-feature

gene

CDS

/codon_start=1

/evidence-not_experimental

/protein_id="AAB70357.1"

/db_xref="PID:g2394471"

/db_xref="GI:2394471"

/translation="MPEKSLLEVPPEVSGPFDHYITTYMTLNTISQSPVCLVKT

ITPNSYCVRPKGLINTGDSKQIVMQLPEKVPDPKQKHEFLVQSCVAPSDVDE

SVKRVKPEPTLYNRVLTVDNDTSKSNICEICTLEFSSVNOQIPRILOCHSHIC

OSCAKSLAKNCVILPCPCRNFTNVSVALEKNAALQALIDOKLVPLEIAKROO

LSCOLSKLEOVSTDKSIKAVGAOLSNLKGREPOEKDSIYSYASIRETINALE

NEAGIKLHTIDMNVSKNALVLAELSEIKRONIKIETKISMEDVDELATRSITDS

GTECAVVPANLVDALIPPEIDKRALSMITTFPSNSTV"

/gene="F42G2.3"

/complement(10285..10355,10407..11300,11351..11465)

/gene="F42G2.3"

/codon_start=1

/evidence-not_experimental

/protein_id="AAB70355.1"

/db_xref="PID:g2394469"

/db_xref="GI:2394469"

/translation="MNSKPLSYPSISITLOHLKPKKFLSQOCCRTTISDKSPLY

LKSKINENSKIKGTYTELVNINENSKPRAPFPDVNSQSIDLEMYVPOEOEEL

VDDRNDPTSTNEORISRADIKRNPSSYTHYIRSYGNHSEQLQYNEKRLHDKML

IDKLGGPPTIKARQFCVENSGLIRIPHEHLIRAKVQNLGILREKRLNATQHLIDK

SLPLSGVILSDPLNANPLDIDFLOTAEMLLYCYEGLISGELIVRHRVYINAYLD

ATFLFEDMHEGREGVTHFSFARNENGINIDVDKIRINYNVEGTAEQNLTHINN

FTDIVSHKVDVTLKXVQREV"

/gene="F42G2.2"

/complement(117255..20720)

/gene="F42G2.2"

/codon_start=1

/evidence-not_experimental

/protein_id="AAB70354.1"

/db_xref="PID:g2394468"

/db_xref="GI:2394468"

/translation="MSYAYNLCOATNMGILTRLDVSGCHGFSQRTDNGKTLTFED

KPVRRVHVDGDTSTYRITTKRKHVOGKTVYDWCEDIHILLNATLALNFEFLT

YVDINSQVKAADRVYEQIKLSQKQKGVYLYLQIDVSLRLVDFDTTLT

ELNINISIIHKPEIIEFGDIYEDQNMADVILHRLNFVSIPEHFSCHESAVIEW

KISADIDOKLIKVDPEADDERIRIVKKSLEDPVTKLKLKRTYPRKRVLYMKNRE

LRCRTGMWFIYHPDYSDSLRNSKREAIAGSOLSLSTSGSTKLVNSYVMGN

OKTALTELELOEYKPEMSQHSYIKCFEFTVPLRLSVEKFTKLTJNAQYVGL

LKYLDPVSLKRIEFGSKRVNENKPMKHILNDTVNCKEHNKNAKELTHTVVK

CMLVYLHFSVVDIILDSILQDVVYIKDALDPTQFOFAQVNSDINETIFDDAP

FRYEPRRIEDSSSLPNIWFEPYDVEEDLIIRYQERKLFEMTRIKRQAL"

/complement(17255..20720)

/gene="F42G2.6"

/complement(17255..17415,17644..18898,18958..19040,

19094..19147,19772..20720)

/gene="F42G2.6"

/note="contains similarity to homeobox domains; coded for

by C. elegans cdna yk74c12.5; coded for by C. elegans cdna

yk51c3.5; coded for by C. elegans cdna yk20c12.5; coded

for by C. elegans cdna yk20c12.3; coded for by C. elegans

cdna yk51c3.3; coded for by C. elegans cdna yk74c12.3"

/codon_start=1

/protein_id="AAB70358.1"

/db_xref="PID:g2394472"

/db_xref="GI:2394472"

/translation="MPTSSPTPIYEDLDLEITYQDCYCHRELYFCNKKMGAGLTFK

NSGINSIRVQDFCCVYKQDEIDSESEVVKFGTGLDRVQVNNCEQHLSPLS

KFCBHYGMIDFENNSLNDTVADVIDSGKDKGCGCFHEDVYLTINIFPKSH

RDEYGOITICDIDNMHTREHGKMSIFESQGFERYMNRVNAOCFKLAKQADVTEY

EVADRTOGCEPESGIESLNPONKLVLPEDLKALIAVPEASLPVCRSPALFPG

KILAEIVAKFEIRFPDPKRIIMFLLIOTLTIRLRDNRKANSNEDIFTSOYQTKL

EYKTKYISSDCKIKSEKIHDELELEKRRARRARRRANSNEDIFTSOYQTKL

TKAGTSGYFLRNKHEIFKTLVTVGSDGVKRCVWTEKLTSEYQIIDQETWIKFEN

SRVAIDAEAREQVHRQKRLNSNKEKEEENDEOKLFEMAELEIONVFIPLNSSD

OISRAATLAEEYCTASNIKSRVNSVLSAISTQQRKIKYKVPNGLVYCGEIL
 TSECKEKVNIDFEPRKINTSLYICDNKFTHEALDELSDQKFGFTMDNGALFG
 TLGNTDIDVARKFSVDLPKRHGGQSALRPARLREKRHYVRKVELAVONFITND
 KVNNGIYIAGSADFKNDLNASDFKNDLAFKIVKVDVSYGGENFOALIASLITL
 GNVKFIQEKILIGYFEEISQDTRICYGEDTLKALELAGVELIYENLEVRWVL
 KDSNGAELIISTROODANRDRMDKGTGOEMEVVQSSEFLVIAEHYKDFGTTLEF
 VSDRSTEGNQVKKFGGIGILRKVNFEOLEVEDDDEYTD*

BASE COUNT 597 a 802 c 616 g 611 t

ORIGIN

Query Match 55.3%: Score 21; DB 8; Length 2626;
 Best Local Similarity 73.0%: Pred. No. 76;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AGGATCAGACATGCTAGGCTTCTTACCCATCA 38
 ||| ||||| || ||| ||||| ||||| ||
 Db 977 AGGTCAGAAAAGACATACGTATCTTCAGCCACCA 941

RESULT 15
 AC004897 90731 bp DNA PRI 14-JAN-1999
 LOCUS AC004897 Homo sapiens clone D0811N16, complete sequence.
 DEFINITION AC004897
 ACCESSION AC004897
 NID 94156184
 VERSION AC004897.1 GI:4156184
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE 1 (bases 1 to 90731)
 JOURNAL Waterston, R.H.
 REFERENCE The sequence of Homo sapiens clone
 AUTHORS Unpublished
 TITLE 2 (bases 1 to 90731)
 JOURNAL Waterston, R.H.
 REFERENCE Direct Submission
 AUTHORS Submitted (12-JUN-1998) Genome Sequencing Center, Washington
 TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 JOURNAL MO 63108, USA

REFERENCE 3 (bases 1 to 90731)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (14-JAN-1999) Genome Sequencing Center, Washington
 TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 JOURNAL MO 63108, USA

COMMENT On Jan 14, 1999 this sequence version replaced gi:3213100.
 FEATURES
 LOCATION/Qualifiers
 1. 90731
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="D0811N16"

BASE COUNT 24311 a 21133 c 20819 g 24468 t

ORIGIN

Query Match 55.3%: Score 21; DB 11; Length 90731;
 Best Local Similarity 73.0%: Pred. No. 64;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

vY 1 AAGATCAGACATGCTAGGCTTCTTACCCATC 37
 ||| ||| ||||| ||| ||||| ||||| ||
 Db 72100 AAGGACAGACATGCTAGGCTTCTTAAATC 72136

Search completed: September 13, 1999, 16:20:17
 Job time: 4529 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:22:01 ; Search time 133.08 Seconds

(without alignments)
71.440 Million cell updates/sec

Title: US-09-325-095-2425_SEQ

Perfect score: 38
Sequence: 1 AAGGATCAGAACATGCGCTTTCCTTTAGCCATCA 38

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	52.6	658	1	Hepatocyte growth
2	19.4	51.1	4000	1	Mannose-1-phosphat
3	19.4	51.1	1078	1	N. excelsior water
4	19.2	50.5	1863	1	Human FRX1 gene. R
5	19.2	50.0	3412	1	Human p110 CDNA. R
6	19	50.0	3207	1	p110 CDNA. Recombi
7	19	50.0	3498	1	Pituitary 3-kinase 11
8	18.8	49.5	1332	1	Enterococcus faeca
9	18.6	48.9	1731	1	Human papilloma vi
10	18.6	48.9	900	1	Staphylococcus aur
11	18.6	48.9	58407	1	Methanococcus jann
12	18.6	48.9	3276	1	Human secreted pro
13	18.6	48.9	2933	1	Human secreted pro
14	18.6	48.9	6234	1	Staphylococcus aur
15	18.4	48.4	12923	1	Sequence of human
16	18.4	48.4	5186	1	E. coli PPC gene
17	18.2	47.9	134525	1	Total base sequenc
18	18.2	47.9	9100	1	H. influenzae dete
19	18.2	47.9	2088	1	Coffee-fruit speci
20	18.2	47.9	110000	1	Continuation (9 of
21	18.2	47.4	1581	1	Catalase gene. Hum
22	18	47.4	320	1	Human genome fragm
23	18	47.4	3159	1	Arg-gingipain-1 ge
24	18	47.4	7266	1	Arg-gingipain-2 ge
25	18	47.4	3094	1	Porphyrinomas ging
26	18	47.4	3071	1	Human chromosome-1
27	18	47.4	8640	1	P. gingivalis haem
28	18	47.4	1296	1	H. pylori cell env
29	18	47.4	7265	1	Arg-gingipain high
30	18	47.4	3159	1	DNA encoding Arg s
31	18	47.4	2282	1	Nucleotide sequenc
32	18	47.4	8640	1	Haemagglutinin pro
33	18	47.4	373	1	EST clone C2598. N
34	18	47.4	4403	1	Staphylococcus aur
35	18	47.4	6482	1	Staphylococcus aur
36	18	47.4	8519	1	Enterococcus faeca
37	18	47.4	6422	1	Polynucleotide seq
38	18	47.4	4546	1	Human kidney amino
39	18	47.4	2264	1	Catalase gene of h
40	17.8	46.8	540	1	Human flap gene. N
41	17.8	46.8	540	1	Sequence encoding
42	17.8	46.8	14985	1	Human box-dependen
43	17.8	46.8	5035	1	Full-length PANCA

ALIGNMENTS

C 44	17.8	46.8	14985	1	V15693	Human Bin1 gene, B
C 45	17.8	46.8	387	1	V36214	cDNA sequence of t

RESULT 1
T78411
ID T78411; standard; cDNA; 658 BP.
AC T78411;
DE 15-OCT-1997 (first entry)
HE Hepatocyte growth factor variant 2.2kb HGF/NK1 terminal portion cDNA.
KW HGF; mitogenesis; proliferative disorder; Kringle domain; inhibitor;
OS cancer; transplant; toxin; chemotherapeutic; human; ss.
FH Homo sapiens.
FH Key
FT cds
FT location/Qualifiers
FT 1..39
FT /*tag= a
FT /*note= "Terminal 2.2kb HGF/NK1 fragment"
FT primer_bind
FT 44..62
FT /*tag= b
FT primer_bind
FT 621..640
FT /*tag= c
FT msc_feature
FT 522..528
FT /*tag= d
FT /*note= "Undetermined sequence"
FN W09640914-A1.
PD 19-DEC-1996.
PR 07-JUN-1996; U08698.
PR 07-JUN-1995; US-484841.
PR (AARO/) AARONSON S A.
PR (BOT/) BOTTAIO D P.
PR (CHAN/) CHAN A M L.
PR (CIOC/) CIOCE V.
PR (RUB/) RUBIN J S.
PR (STAH/) STAHL S J.
PR (WING/) WINGFIELD P T.
PI Aaronson SA, Bottario DP, Chan AML, Cioce V, Rubin JS;
PI Stahl SJ, Wingfield PT;
DR WPI: 97-099934/09.
DR P-BSDB: W23714.
PT Truncated hepatocyte growth factor (HGF) variants - used to inhibit
PT HGF-induced mitogenesis, useful in treating proliferative disorders
PS Example 2; Fig 10; 89pp; English.
CC The present sequence encodes the terminal portion of the coding sequence
CC and part of the 3'-untranslated region of the truncated hepatocyte
CC growth factor (HGF) 2.2kb HGF/NK1. HGF/NK1 is a substantially pure
CC truncated HGF variant comprising the N-terminal and first kringle domain
CC of HGF. HGF/NK1 may be used in cell culture methods to stimulate ex vivo
CC mitogenesis in cells expressing the HGF, such cells can be transplanted
CC into a subject in need of such cells. HGF/NK2, which is a truncated form
CC of HGF comprising the N-terminal and the first two kringle domains, is
CC used to inhibit HGF-induced mitogenesis in cells expressing the HGF
CC receptor (HGF4), and it is useful in treating hyper-proliferative
CC disorders including cancer and non-malignant disorders. Both HGF/NK1
CC and HGF/NK2 are non-toxic and bind specifically to HGFr and may
CC therefore be used as carriers for other compounds, e.g. permitting
CC the evaluation of the usefulness of certain toxins as chemotherapeutic
CC agents.
SQ Sequence 658 BP; 216 A; 126 C; 126 G; 180 T;

Query Match
Best local similarity 52.6%; Score 20; DB 1; Length 658;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 AAGGATCAGAACATGCGCTTTCCTTTAGCCAT 36
DB 358 AAGGATCAGAACATGCGCTTTCCTTTAGCCAT 393

RESULT 2

```

T91902
ID T91902 standard: DNA: 4000 BP.
AC T91902:
DT 07-MAY-1998 (first entry)
DE Mannose-1-phosphate transferase gene MNM4.
KW Mannose-1-phosphate transfer; MNM4 gene; enzyme; yeast; regulatory gene;
KW human; high mannose type neutral saccharide chain; ss.
OS Saccharomyces cerevisiae.
FH Key
FT Location/Qualifiers
FT CDS
FT 418..3953
FT /*tag= a
FT /product= MNM4

PN J09266792-A.
PD 14-OCT-1997.
PE 29-MAR-1996; 075667.
PR 29-MAR-1996; JP-075667.
PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
PR MPI; 97-553460/51.
DR P-PSDB; M30763.
PT Positive regulatory gene of mannose-1-phosphate transfer in yeast -
PT useful for high mannose type neutral saccharide chain production
PS Claim 2; Page 14-17; 23pp; Japanese.
CC This sequence represents the gene of the invention, designated MNM4,
CC encoding a protein which positively regulates mannose-1-phosphate
CC transfer in yeast. The gene is useful for the preparation of human high
CC mannose type neutral saccharide chain.
SQ Sequence 4000 BP; 1453 A; 750 C; 770 G; 1027 T;

Query Match 51.1%; Score 19.4; DB 1; Length 4000;
Best Local Similarity 70.3%; Pred. No. 20;
Matches 26; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

UY 1 AAGCATCAGAACATGCCAGCGCTTCCTTATCCATC 37
DB 1861 AATGATCAGAACATTCCAAAGCTTACTTAAGCCAAC 1897

RESULT 3
V56043/C
ID V56043 standard: cDNA: 1078 BP.
AC V56043:
DT 09-DEC-1998 (first entry)
DE N. excelsior water channel protein, Nemi1p encoding cDNA.
KW Nicotiana: cell-membrane localised water channel protein; plant;
KW water potential; plant activity; Nemi1p; Nemi2p; Nemi3p;
KW water stress resistance; ds.
OS Nicotiana excelsior.
FH Key
FT Location/Qualifiers
FT CDS
FT 34..897
FT /*tag= a
FT /product= "cell-membrane localised water channel
FT protein Nemi1p."

PN J102239883-A.
PD 02-SEP-1998.
PE 24-FEB-1997; 039163.
PR 24-FEB-1997; JP-039163.
PA (NISB ) JAPAN TOBACCO INC.
PR MPI; 98-524287/45.
DR P-PSDB; W60568.
PT Improving water potential and maintaining plant activity - comprises
PT introducing plant cell membrane-localised water channel protein gene
PT into plant, produces plants resistant to water stress
PS Examples; Page 6; 11pp; Japanese.
CC This cDNA encodes a Nicotiana excelsior cell-membrane localised water
CC channel protein, Nemi1p. This can be used in a method for improving
CC water potential and maintaining plant activity. The method comprises
CC introducing a plant cell membrane-localised water channel protein gene,
CC especially the N. excelsior Nemi1p, Nemi2p and Nemi3p genes, into a
CC plant. The process can be used to produce plants with resistance to
CC water stress.
SQ Sequence 1078 BP; 261 A; 228 C; 247 G; 342 T;

```

```

Query Match Similarity      51.1%; Score 19.4; DB 1; Length 1078;
Best Local Similarity      70.3%; Pred. No. 15;
Matches      26; Conservative      0; Mismatches      11; Indels      0; Gaps      0

OY      2 AGAGTCAGAACAAATGCTTAGGCTTCTTAGGCATCA 38
      |||| | | | | | | | | | | | | | | | | | |
DB      660 AGCAACATGAGAACTCTAGCATTTCTTGGCATCA 624

RESULT      4
ID      T47705 standard; cDNA; 1863 BP.
AC      T47705;
DT      20-MAY-1997 (first entry)
DE      Human FRX1 gene.
KW      FRX1; FRX2; FRX1; Fragile X syndrome; mental retardation;
OS      diagnosis; transgenic mouse; knockout mouse; ds.
FH      Key
FH      primer_bind      Location/Qualifiers
FT      complement (1777..1798)
FT      /*tag= a
FT      /note= "primer used in FRX1 gene chromosome
FT      mapping"
PN      WO9638467-A1.
PD      05-DEC-1996.
PR      31-MAY-1996; U08853.
PR      31-MAY-1995; US-455073.
PA      (TYPE=) UNIV PENNSYLVANIA.
PI      Dreyfuss G, Sloni MC, Zhang Y;
PI      WPI; 97-034303/03.
PI      P-PSDB; 003429.
PT      New fragile X related genes FRX1 and FRX2 - used to develop prods.
PT      For identifying individuals with mental retardation caused by
PT      fragile X syndrome
PS      Claim 2; Page 57-59; 75pp; English.
CC      cDNA clones (T47705 and T47706) respectively code for novel human
CC      proteins FRX1 (W09429) and FRX2 (W09430), which are related to FMRI,
CC      the protein associated with Fragile X syndrome. To obtain the human
CC      FRX1 cDNA, a Xenopus laevis ovary cDNA library was screened with
CC      human FMRI cDNA to obtain FMRI and FRX1 (for FMRI unreacting
CC      relative) cDNAs. The Xenopus FRX1 cDNA was then used to isolate
CC      human FRX1 cDNA from a HeLa cell library. The FRX1 gene is
CC      autosomal (unlike FMRI) and is located at 12q13. The FRX1 and FRX2
CC      genes can be used to produce FRX1 and FRX2 recombinant polypeptides,
CC      to develop FRX1 and FRX2 knockout mice, and to develop methods of
CC      screening individuals for FMRI deficiency.
SO      Sequence      1863 BP;      620 A;      320 C;      470 G;      453 T;

Query Match      50.5%; Score 19.2; DB 1; Length 1863;
Best Local Similarity      75.0%; Pred. No. 21;
Matches      24; Conservative      0; Mismatches      8; Indels      0; Gaps      0

OY      1 AAGATCAGAACAAATGCTTAGGCTTCTTAG 32
      ||| | | | | | | | | | | | | | | | | |
DB      1050 AAGAACCTGACACATTTCCAAATGCTTCTTAG 1019

RESULT      5
ID      Q51156 standard; cDNA; 3412 BP.
AC      Q51156;
DT      12-APR-1994 (first entry)
DE      Human p110 cDNA.
KW      Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW      antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW      platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW      ds.
OS      Human.
FH      Key
FH      Location/Qualifiers
FT      cds      1..3207
FT      /*tag= a

```

FT /note="PI3- kinase p110"
 PN M09321328-A.
 PD 28-OCT-1993.
 PR 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI ParkerPJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 DR P-PSDB: R43342.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Claim 7; Fig 16; 146pp; English.
 CC Southern blot analysis was performed using a bovine cDNA probe contg.
 CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
 CC from a cDNA library constructed from mRNA isolated from the human
 CC cell line K562. Positive clones were sequenced to give the human
 CC PI3 kinase p110 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p110 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to
 CC provide a protein with PI3 kinase activity, and is useful for
 CC screening for (anti)agonists of PI3 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC prophylaxis or therapy. Platelet or neutrophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051155 and 057522-3.
 SO Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 50.0%; Score 19; DB 1; Length 3412;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 AGCGTTCTTTAGCCATCA 38
 ||||||||||||||||
 DB 878 AGCGTTCTTTAGCCATCA 860

RESULT 6
 ID 051155 standard; cDNA; 3207 BP.
 AC 051155.
 DT 12-APR-1994 (first entry)
 DE PI10 cDNA.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KM ss.
 FH Key Location/Qualifiers
 FT 1..3207
 FT cds /*tag= a
 FT /*note="PI3- kinase p110"
 PN M09321328-A.
 PD 28-OCT-1993.
 PR 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI ParkerPJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 DR P-PSDB: R43341.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Claim 7; Fig 9; 146pp; English.
 CC Southern blot analysis was performed using a bovine cDNA probe contg.
 CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
 CC from a cDNA library constructed from mRNA isolated from the human
 CC cell line K562. Positive clones were sequenced to give the human
 CC PI3 kinase p110 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p110 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for
 CC screening for (anti)agonists of PI3 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC prophylaxis or therapy. Platelet or neutrophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051155, 059012-23 and 057522-3.
 SO Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match 50.0%; Score 19; DB 1; Length 3207;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 AGCGTTCTTTAGCCATCA 38
 ||||||||||||||||
 DB 878 AGCGTTCTTTAGCCATCA 860

RESULT 7
 ID 057012 standard; cDNA to mRNA; 3498 BP.
 AC 057012.
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase 110 KD catalytic subunit cDNA.
 KW 110 KD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transfection; Schizosaccharomyces pombe; nut promoter; thiamine;
 KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Bos taurus.
 FH Key Location/Qualifiers
 FT 1..3207
 FT cds /*tag= a
 FT /*product= p110
 PN M09403609-A.
 PD 17-FEB-1994.
 PR 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 DR P-PSDB: R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure; Fig 1; 71pp; English.
 CC This sequence encodes the 110 KD catalytic subunit of the phosphatidylinositol (Ptdins) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nut promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the ptdins
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nut promoter functions and the ptdins 3-kinase
 CC catalytic subunit is induced. Ptdins activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SO Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 50.0%; Score 19; DB 1; Length 3498;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 AGCGTTCTTTAGCCATCA 38
 ||||||||||||||||
 DB 878 AGCGTTCTTTAGCCATCA 860

RESULT 8
 ID X13394 standard; DNA; 1332 BP.

```

AC X13394;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:457.
KM Enterococcus faecalis; contig; detection; Enterococcal infection;
OS vaccine; attenuation; computer readable medium; ds.
PN MO9850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; 008985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA.
DR WPI; 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1: Page 1690-1691; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism. In vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 1332 BP; 479 A; 190 C; 227 G; 431 T.

Query Match 49.5%; Score 18.6; DB 1; Length 1332;
Best Local Similarity 76.7%; Pred. No. 29;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 9 GAACAATGCTAGCTTCTTACCATCA 38
DB 496 GATACATGCTATGCTTTTTCAGATCA 467

RESULT 9
ID N80161 standard; DNA; 1731 BP.
AC N80161;
DT 19-OCT-1990 (first entry)
DE Human papilloma virus 18 protein
KM Human papilloma virus; HPV18; malignant genital tumours; ss.
FH Key Location/Qualifiers
FT cds 99..572
FT /tag= a
FT /label=HPV18 E6 protein
FT cds 584..898
FT /tag= b
FT /label=HPV 18 E7 protein
FT cds 908..729
FT /tag= c
FT /label=HPV 18 E1 protein
PN DE3625257-A.
PD 04-FEB-1988.
PF 25-JUL-1986; 625257.
PR 23-JUL-1986; DE-624786.
PA (BEHN) Behringwerke AG.
PI Becker H, Meyer-Dulheuer K;
DR WPI; 88-037095/06.
PT Expression prods. of human papilloma viruses type 16 and 18 -
PT used for the prodn. of diagnostics for HPV16 and HPV18 infections

```

```

PS Disclosure; German.
CC The expression prods. of HPV18, contained in this sequence, are used to
CC produce antibodies to detect these. HPV16 expression prods are also used
CC for antibody prods. See also N80155-59.
SQ Sequence 1731 BP; 610 A; 313 C; 403 G; 405 T;

Query Match 48.9%; Score 18.6; DB 1; Length 1731;
Best Local Similarity 84.0%; Pred. No. 38;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 9 GAACAATGCTAGCTTCTTACCATCA 33
DB 1117 GATACATGCTATGCTTCTTACCATCA 1093

RESULT 10
ID T88772 standard; DNA; 900 BP.
AC T88772;
DT 23-MAR-1998 (first entry)
DE Staphylococcus aureus ERA encoding DNA.
KM GTP binding protein family; ERA; immune response; infection;
OS diagnosis; antibody; respiratory tract; ss.
PN MO9730149-A1.
PD 21-AUG-1997.
PF 19-FEB-1997; 002547.
PR 20-FEB-1996; US-011888.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PI Black WT, Burnham MKR, Hodgson JE, Knowles DJC,
PI Nicholas RO, Pratt DM, Reichard RW, Rosenberg M,
PI Ward DM;
DR WPI; 97-425028/39.
PT P-PSDB; W31905.
PT Staphylococcus aureus ERA gene - used to induce immune response
PT against diseases related to S. aureus infection
PS Claim 1: Fig 1; 46pp; English.
CC The present sequence encodes a novel isolated GTP binding protein
CC referred to as ERA. ERA can be used for diagnosing a disease related
CC to ERA expression. The ERA protein can be used for inducing an
CC immunological response in a mammal, to produce antibodies to protect
CC the animal from disease. The disease is an infection by Staphylococcus
CC aureus (or another bacteria), and include infections of the upper
CC respiratory tract (e.g. otitis media, acute epiglottitis, thyroiditis),
CC lower respiratory tract (e.g. lung abscess), cardiac (e.g. infective
CC endocarditis), gastrointestinal tract (e.g. secretory diarrhoea, splenic
CC abscess), central nervous system (e.g. cerebral abscess), eye (e.g.
CC conjunctivitis), kidney and urinary tract (e.g. epididymitis, toxic
CC shock syndrome), skin (e.g. folliculitis, cellulitis, wound infection),
CC and bone and joint (e.g. septic arthritis).
SQ Sequence 900 BP; 348 A; 118 C; 179 G; 255 T;

Query Match 48.9%; Score 18.6; DB 1; Length 900;
Best Local Similarity 72.7%; Pred. No. 33;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 GATCAGACATGCTAGCTTCTTACCATCA 36
DB 684 GATATGACACATGCTTCTTACCATCA 652

RESULT 11
ID V21210 standard; DNA; 58407 BP.
AC V21210;
DT 10-NOV-1998 (first entry)
DE Methanococcus jannaschii large circular extrachromosomal element.
KM Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KM genome; autotrophic; extrachromosomal element; identification; ds.
OS Methanococcus jannaschii.
PN MO9807830-A2.

```


PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057659.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DE, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferlie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z,
PI WPI: 98-506364/43.
DR P-FSDB: W75006.
DR
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1: Page 519-521; 721pp: English.
CC This sequence represents a nucleic acid molecule designated Gene 155 from
CC the human cDNA clone H1XFJ55 (deposited as clone ATCC 97903 and ATCC
CC 209049) which encodes a secreted human protein. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. V59502) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see V59511 for described
CC uses).
CC
CC Sequence 3276 BP; 811 A; 783 C; 928 G; 747 T;
CC

Query Match 48.9%; Score 18.6; DB 1; Length 3276;
Best Local Similarity 72.7%; Pred. No.44;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 AAGATCAGACATGCTAGGCTTCTTACG 33
DB 1406 ATGATCTGACAGATCTCTAAGCGTCTGTACG 1374

RESULT 13
V59665/c
ID V59665 standard; DNA: 2933 BP.
AC V59665;
DT 19-JAN-1999 (first entry)
DE Human secreted protein gene 155 clone H1XFJ55.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; osteoclast; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN MO9839448-A2.
PD 11-SEP-1998
PR 06-MAR-1998; U04493.
PR 02-OCT-1997; US-061060.
PR 07-MAR-1997; US-038621.
PR 07-MAR-1997; US-040161.

PR 07-MAR-1997; US-040162.
PR 07-MAR-1997; US-040163.
PR 07-MAR-1997; US-040333.
PR 07-MAR-1997; US-040334.
PR 07-MAR-1997; US-040336.
PR 07-MAR-1997; US-040626.
PR 11-APR-1997; US-043311.
PR 11-APR-1997; US-043312.
PR 11-APR-1997; US-043313.
PR 11-APR-1997; US-043314.
PR 11-APR-1997; US-043315.
PR 11-APR-1997; US-043368.
PR 11-APR-1997; US-043369.
PR 11-APR-1997; US-043376.
PR 11-APR-1997; US-043377.
PR 11-APR-1997; US-043378.
PR 11-APR-1997; US-043380.
PR 11-APR-1997; US-043659.
PR 11-APR-1997; US-043670.
PR 11-APR-1997; US-043671.
PR 11-APR-1997; US-043672.
PR 11-APR-1997; US-043674.
PR 23-MAY-1997; US-047492.
PR 23-MAY-1997; US-047493.
PR 23-MAY-1997; US-047501.
PR 23-MAY-1997; US-047502.
PR 23-MAY-1997; US-047503.
PR 23-MAY-1997; US-047581.
PR 23-MAY-1997; US-047582.
PR 23-MAY-1997; US-047583.
PR 23-MAY-1997; US-047584.
PR 23-MAY-1997; US-047585.
PR 23-MAY-1997; US-047586.
PR 23-MAY-1997; US-047587.
PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 13-JUN-1997; US-049610.
PR 08-JUL-1997; US-051826.
PR 16-JUL-1997; US-052874.
PR 18-AUG-1997; US-055724.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056633.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.

```

PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056883.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056885.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056890.
PR 22-AUG-1997; US-056891.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057659.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarek DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrite AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kryw M, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben EM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z,
DR WPI: 98-506364/43.
DR P-PSDB: W74883.
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1: Page 391-393; 721pp; English.
CC This sequence represents a nucleic acid molecule designated Gene 155 from
CC the human cDNA clone HMXFJ55 (deposited as clone ATCC 97903 and ATCC
CC 209049) which encodes a secreted human protein. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. V59502) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W4731-W5026) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see V59511 for described
CC sequence).
SQ Sequence 2933 BP; 726 A; 715 C; 833 G; 651 T;

Query Match 48.9%; Score 18.6; DB 1; Length 2933;
Best Local Similarity 72.7%; Pred. No. 43;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 AAGGATCAGAACATGCTAGCCTTCTTACG 33
DB 1409 ATGATCTGAGAGATCTTAAGCGCTGTAGC 1377

```

```

FT misc-feature 1561..1620
FT /*tag- a
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 3361..3420
FT /*tag- b
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 5161..5220
FT /*tag- c
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA,
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S aureus vaccines
PS Claim 1: Page 1297-1300; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S. aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S. aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S. aureus in a sample. S. aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S. aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 6234 BP; 2215 A; 822 C; 1171 G; 1852 T;

Query Match 48.9%; Score 18.6; DB 1; Length 6234;
Best Local Similarity 72.7%; Pred. No. 50;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 GATCAGAACATGCTAGCCTTCTTACCAT 36
DB 5938 GATATGACACAGCATCTTCGCTTCTTACCAT 5906

```

```

RESULT 14
V74715/c
ID V74715 standard; DNA: 6254 BP.
AC V74715;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #404.
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers

```

```

RESULT 15
N90338
ID N90338 standard; cDNA: 12923 BP.
AC N90338;
DT 29-MAR-1992 (first entry)
DE Sequence of human muscular dystrophy (MD) cDNA.
KW Dystrophin; muscular dystrophy; probe; antibody; diagnosis;
KW prenatal; heterozygote; gene therapy; genetic screening;
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 209..12923
FT /*tag- a

```

PN MO8906286-A.
 PD 13-JUL-1989.
 PF 16-DEC-1988; U04504.
 PR 22-DEC-1987; US-136618.
 PA (CHIL-) CHILDRENS MED CENT.
 PI Kunkel LM, Monaco A, Hoffman EP, Koenig M;
 DR WPI; 89-220587/30.
 P-PSDB; P90373.
 PT Muscular dystrophy gene - used for prepn. of probes, dystrophin
 PT polypeptide and antibodies for diagnosis and therapy of muscular
 PT dystrophy
 PS Disclosure: Fig 5: 68pp; English.
 CC The inventors claim an MD probe comprising a purified ss NA, SQ which
 CC hybridises to at least a part of the MD gene; pure dystrophin (DS)
 CC polypeptide, purified NA encoding DS and antibodies (Ab) to DS. The
 CC probes are equal to or greater than 10b of one of 12 cDNA sequences
 CC deposited as ATCC 58666-57677. The MD gene is human, or a murine Dmd
 CC gene.
 SQ Sequence 12923 BP; 4296 A; 2613 C; 2963 G; 3044 T;

Query Match

48.4%; Score 18.4; DB 1; Length 12923;

Best Local Similarity 69.4%; Pred. No. 72;

Matches 25; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AGGATCAGAACATGCTAGGCTTCTTTAGCCATC 37
 II IIIIIIIII I IIIII II II I
 DB 2688 AGTATCAGAACATCATCATGCTTCTATATATCAGC 2723

Search completed: September 13, 1999, 16:22:05
 Job time: 4025 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:13:58 ; Search time 849.94 Seconds
(without alignments)
88.190 Million cell updates/sec

Title: US-09-325-095-2425.SEQ

Perfect score: 38

Sequence: 1 AAGATCAGACAAATGCCCTAGGCTTCTTAGCCARCA 38

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database: EST:*

1: em_est1:*

2: em_est2:*

3: em_est3:*

4: em_est4:*

5: em_est5:*

6: em_est6:*

7: em_est7:*

8: em_est8:*

9: em_est9:*

10: em_est10:*

11: em_est11:*

12: em_est12:*

13: em_est13:*

14: em_est14:*

15: em_est15:*

16: em_est16:*

17: em_est17:*

18: em_est18:*

19: em_est19:*

20: em_est20:*

21: em_est21:*

22: em_est22:*

23: em_est23:*

24: em_est24:*

25: em_est25:*

26: em_est26:*

27: em_est27:*

28: em_est28:*

29: em_est29:*

30: em_est30:*

31: em_est31:*

32: em_est32:*

33: em_est33:*

34: em_est34:*

35: em_est35:*

36: em_est36:*

37: em_est37:*

38: em_est38:*

39: em_est39:*

40: em_est40:*

41: em_est41:*

42: em_est42:*

43: em_est43:*

44: em_est44:*

45: em_est45:*

46: em_est46:*

47: em_est47:*

48: em_est48:*

49: em_est49:*

50: em_est50:*

51: em_est51:*

52: em_est52:*

53: em_est53:*

54: em_est22:*

55: em_est23:*

56: em_est24:*

57: em_est25:*

58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	21.8	57.4	546	54	HSW010244	A1045394 Homo sapi
C 2	21	55.3	249	37	AA716393	AA716393 zebrafish
C 3	21	55.3	355	39	AA903710	AA903710 Oryzias latipes
C 4	21	55.3	279	44	A1282767	A1282767 Drosophila melanogaster
C 5	20.8	54.7	325	46	A1029617	A1029617 Drosophila melanogaster
C 6	20.6	54.2	371	21	R02605	R02605 Drosophila melanogaster
C 7	20.6	54.2	411	33	AA446728	AA446728 Drosophila melanogaster
C 8	20.6	54.2	484	40	AA922889	AA922889 Drosophila melanogaster
C 9	20.4	53.7	700	54	HSW010166	A1045316 Homo sapi
C 10	20.2	53.2	717	22	R81085	R81085 EST2201
C 11	20.2	53.2	426	24	H84400	H84400 Yf01a10.r1
C 12	20.2	53.2	400	27	AA018084	AA018084 zebrafish
C 13	20.2	52.6	390	29	AA172139	AA172139 Drosophila melanogaster
C 14	20	52.6	307	46	A1425878	A1425878 Drosophila melanogaster
C 15	19.8	52.1	628	38	AA799278	AA799278 Drosophila melanogaster
C 16	19.8	52.1	1006	48	A1573683	A1573683 Drosophila melanogaster
C 17	19.6	51.6	581	25	N58217	N58217 Yf66h06.s1
C 18	19.6	51.6	416	25	N83070	N83070 T9EST2760c1
C 19	19.6	51.6	142	26	W70126	W70126 zebrafish
C 20	19.6	51.6	205	29	AA193522	AA193522 zebrafish
C 21	19.6	51.6	542	30	AA214230	AA214230 zebrafish
C 22	19.6	51.6	330	33	AA426031	AA426031 zebrafish
C 23	19.6	51.6	378	37	AA719187	AA719187 zebrafish
C 24	19.6	51.6	589	38	AA264378	AA264378 Drosophila melanogaster
C 25	19.6	51.6	546	38	AA438766	AA438766 Drosophila melanogaster
C 26	19.6	51.6	790	39	AA843153	AA843153 Drosophila melanogaster
C 27	19.6	51.6	439	42	A1129154	A1129154 Drosophila melanogaster
C 28	19.6	51.6	474	43	A1183972	A1183972 Drosophila melanogaster
C 29	19.6	51.6	429	47	A1512524	A1512524 Drosophila melanogaster
C 30	19.6	51.6	458	49	A1633740	A1633740 Drosophila melanogaster
C 31	19.6	51.6	451	50	AU065351	AU065351 Drosophila melanogaster
C 32	19.4	51.1	336	20	T30439	T30439 EST16732 Hu
C 33	19.4	51.1	325	21	R05870	R05870 yec3b08.r1
C 34	19.4	51.1	629	29	AA141078	AA141078 CR01174.3
C 35	19.4	51.1	1019	35	C23297	C23297 C23297 Jappa
C 36	19.4	51.1	526	38	AA786092	AA786092 zebrafish
C 37	19.4	51.1	646	39	AA820961	AA820961 Drosophila melanogaster
C 38	19.4	51.1	586	40	AA948780	AA948780 Drosophila melanogaster
C 39	19.4	51.1	803	41	A1069025	A1069025 Drosophila melanogaster
C 40	19.4	51.1	423	43	A1209388	A1209388 Drosophila melanogaster
C 41	19.4	51.1	328	45	A1331461	A1331461 Drosophila melanogaster
C 42	19.4	51.1	439	45	A1332069	A1332069 Drosophila melanogaster
C 43	19.4	51.1	721	45	AU004032	AU004032 Drosophila melanogaster
C 44	19.4	51.1	534	46	A1455453	A1455453 Drosophila melanogaster
C 45	19.4	51.1	276	47	A1540728	A1540728 Drosophila melanogaster

ALIGNMENTS

RESULT 1
ID HSW010244/c standard; RNA; EST; 546 BP.
XX
AC A1045394;
XX
SV A1045394.1
XX

NI e1403868
 XX 12-MAR-1999 (Rel. 59, Created)
 DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
 XX
 DE Homo sapiens mRNA; EST DKFZP434D205_r1 (from clone
 DE DKFZP434D205)
 XX
 KW EST: expressed sequence tag.
 XX
 XX Homo sapiens (human)
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
 OC Primates; Catarrhini; Homiidae; Homo.
 XX
 XX [1]
 RP 1-546
 RA Duesterhoeft A., Lauber J., Mewes W., Gassenhuber J., Wiemann S.;
 RT Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
 RL M185, Am Klopferstr. 18a D-82152 Martinsried, GERMANY
 XX
 CC Clone from S. Wiemann, sequenced by Qiagen within the cDNA
 CC sequencing consortium of the German Genome Project
 CC s1 sequence also available
 CC This clone is available at the RZPD in Berlin
 CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
 XX
 XX Key Location/Qualifiers
 FH
 FH source 1..546
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT /clone_id="434 (synonym: htes3). Vector pSport1; host
 FT DH10B; sites NotI + SalI
 FT /dev_stage="adult"
 FT /tissue_type="testis"
 XX
 SO Sequence 546 bp; 156 A; 123 C; 143 G; 124 T; 0 other;

Query Match 57.4%; Score 21.8; DB 54; Length 546;
 Best Local Similarity 78.8%; Pred. No. 18;
 Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 5 ATCAGACAAATGCTAGCGCTTCTTACCATC 37
 Db 118 ATCAGACAAATGCTAGCGCTTCTTACCATC 86

RESULT 2
 AA716393 249 bp mRNA EST 29-DEC-1997
 LOCUS 2962d06.s1 Soares_fetal_heart.NBHL19W Homo sapiens cDNA clone
 DEFINITION IMAGE:397931.3', mRNA sequence.
 ACCESSION AA716393
 MID 92728667
 VERSION AA716393.1 GI:2728667
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 249)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Martin, J., Moore, B., Kucaba, T., Lacy, M., Lennon, G., Marra, M.,
 Theising, B., White, F., Wylie, T., Waterston, R. and Wilson, R.
 TITLE WashU-NCI human EST project
 JOURNAL Unpublished (1997)
 COMMENT On May 8, 1995 this sequence version replaced gi:801246.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: east@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 135.
 Location/Qualifiers
 1..249
 /organism="Homo sapiens"
 /db_xref="GDB:1305754"
 /db_xref="taxon:9606"
 /clone_id="IMAGE:397931"
 /clone_id="Soares_fetal_heart_NBHL19W"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a
 modified polylinker; site_1: Not I; site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTACCAATCTGAGAGGAGGCGGCGCATCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library was constructed by
 M. Fatima Bonaldo. This library was constructed from the
 same fetus as the fetal lung library; Soares fetal lung
 NBHL19W."

BASE COUNT 68 a 62 c 47 g 72 t
 ORIGIN

Query Match 55.3%; Score 21; DB 37; Length 249;
 Best Local Similarity 73.0%; Pred. No. 37;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AGATCAGACAAATGCTAGCGCTTCTTACCATCA 38
 Db 8 AGATCCTTACCATGCTAGCGCTTCTTACCATCA 44

RESULT 3
 AA903710 355 bp mRNA EST 08-APR-1998
 LOCUS 060a11.s1 NCI-CGAP_G04 Homo sapiens cDNA clone IMAGE:151832.3'
 DEFINITION similar to contains Alu repetitive element; mRNA sequence.
 ACCESSION AA903710
 MID 93038833
 VERSION AA903710.1 GI:3038833
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 355)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgaf.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1407050.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bdrp/image/image.html

Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 262.

FEATURES
SOURCE
Location/Qualifiers

1..335
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1518332"
/clone_1lb="NCI-CGAP_GC4"
/tissue="type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"

BASE COUNT
ORIGIN
98 a 83 c 76 g 98 t

Query Match 55.3%; Score 21; DB 39; Length 355;
Best Local Similarity 73.0%; Pred. No. 38;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 2 AGGATCAGACATGCTAGGCTTCTTTAGCCATCA 38
||||| ||||||| ||| |||
Db 7 AGGATCTTACCATGCTAGGCTTCTTTAGCCATCA 43

RESULT 4

LOCUS A1282767 279 bp mRNA EST 21-DEC-1998
DEFINITION q184b06.x1 NCI-CGAP_C014 Homo sapiens CDNA clone IMAGE:1961939 3'
similar to contains Alu repetitive element; contains element LTR6
repetitive element ; mRNA sequence.
A1282767
NID g3821000
VERSION A1282767.1 GI:3921000
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 279)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 8, 1995 this sequence version replaced gi:800256.

JOURNAL
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html

FEATURES
SOURCE
Insert Length: 1300 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 237.
Location/Qualifiers
1..279

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1961939"
/clone_1lb="NCI-CGAP_C014"
/tissue="type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-Sport6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.7 kb. Life Technologies catalog #:
11531-019"

BASE COUNT
ORIGIN
74 a 71 c 55 g 79 t

Query Match 55.3%; Score 21; DB 44; Length 279;
Best Local Similarity 73.0%; Pred. No. 37;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 2 AGGATCAGACATGCTAGGCTTCTTTAGCCATCA 38
||||| ||||||| ||| |||
Db 7 AGGATCTTACCATGCTAGGCTTCTTTAGCCATCA 43

RESULT 5
LOCUS A1029617/c 325 bp mRNA EST 09-MAR-1999
DEFINITION UI-R-C0-1Y-e-09-0-UI.s1 UI-R-C0 Rattus norvegicus CDNA clone
UI-R-C0-1Y-e-09-0-UI 3', mRNA sequence.
A1029617
ACCESSION g4289791
VERSION A1029617.1 GI:4299791
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 325)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL
MEDLINE
COMMENT On Jun 22, 1998 this sequence version replaced gi:3247443.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mesares@blue.weeg.uiowa.edu
The sequence tag present in the CDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult 12-Day-Embryo library. CDNA Library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics
Seq primer: M13 Forward.
Location/Qualifiers

FEATURES
SOURCE

1..325
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C0-1Y-e-09-0-UI"
/clone_1lb="UI-R-C0"
/dev_stage="adult"
/lab_host="DH10B" (Life Technologies)
/note="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
library is a subtracted library derived from the UI-R-A1
and UI-R-E1 libraries. The UI-R-A1 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, and muscle. The UI-R-E1

library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-pT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C0) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had been derived was used as a diver in a hybridization with the pooled UI-R-A1 and UI-R-E1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10 bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996).

Query Match	Similarity	Score	DB	Length
Best Local	Similarity 78.1%	Pred. No. 46		
Matches	25; Conservative	0; Mismatches	7; Indels	0; Gaps
0Y	5 ATCAGACATGCTTAGGCTTCTTTAGGCAT	36		
Db	189 ATCAGCGAGTCCCTGTGCTTTGTTAGGCAT	158		

RESULT	6			
LOCUS	R02605	371 bp	MRNA	EST
DEFINITION	ye16a05.r1 Soares fetal liver spleen INTLS Homo sapiens cDNA clone			
IMAGE:	123634 5', mRNA sequence.			

ACCESSION	R02605
NID	g752341
VERSION	R02605.1
KEYWORDS	GI:752341
SOURCE	EST.
ORGANISM	human.
	Homo sapiens

REFERENCE
AUTHORS
Eutheria; Primates; Carnivora; Hominoidea; Homo.
1 (bases 1 to 371)
Hillier, L., Clark, N., Dubouche, T., Elliston, K., Hawkins, M.,

TREVASKIS, E., WATERSTON, R., WILLIAMSON, A., WOHLDMANN, P. and WILSON, R.
The WASHU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 751
High quality sequence stops: 257 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 751 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 257.

```

FEATURES
source
location/Qualifiers
1. .371
/organism="Homo sapiens"
/db_xref="GDB:476179"
/db_xref="taxon:9606"
/clone="IMAGE:123634"
/clone.lib="Scares fetal liver spleen INFUS

```

```

/sex="male"
/deg_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pTV73D (Pharmacia)
with a modified polylinker. Site.1: Pac I; Site.2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dt) primer
[5' AACGTGACGACATTAAATTAAAGATCTTTTATTTTATTTT 3'],
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTV73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

```

Query Match	54.2%	Score 20.6	DB 21	length 371
Best Local Similarity	74.3%	Pred. No. 56		
Matches 26	Conservative 0	Mismatches 9	Indels 0	Gaps 0

QY	2	AGATCAGACATGCGCTTCTTTAGCCAT	36
Db	117	AAGAGACAACATATATAGGTTCTTTAGGAT	151

RESULT	7			
AA446728	AAA46728	411 bp	mrna	EST 03-JUN-1997
LOCUS				
DEFINITION	Z884163.f1 Soares fetalus_Nb2HF8.9w Homo sapiens cDNA clone IMAGE:783677 5', mRNA sequence.			

NID	G2159393
VERSION	A4446728.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Crustacea; Vertebrata; Mammalia;
Eutheria; Primates; Carnivora; Hominoidea; Homo.
1 (bases 1 to 411)
Hillier, L., Allen, M., Bowles, L., Dubouque, T., Gelsel, G., Jost, S.,
Kucaba, T., Laey, M., Le, N., Lennon, G., Marra, M., Martin, J.,
Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,

TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT On Jan 25, 1995 this sequence version replaced q1:637779

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: estlevatson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28mb3 rev2 Et from Amersham
 High quality sequence stop: 343.
 Location/Qualifiers
 1..411

/organism="Homo sapiens"
 /db_xref="GDB:5981295"
 /db_xref="taxon:9606"
 /map="5"
 /clone="IMAGE:783677"
 /clone_1lb="Scores_total_fetus_Nb2HF8_9w"
 /dev_stage="8-9 weeks"
 /lab_host="DH10B"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand
 (total) fetus material with a Not I - oligo(dT) primer [5'
 TGTTCCATCTGAAATGGAGCGCCGCTTAATTTTTTTTTTTT 3']
 Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Palma Bonaldo. "

BASE COUNT 115 a 57 c 63 g 176 t
 ORIGIN

Query Match 54.2%; Score 20.6; DB 33; Length 411;
 Best Local Similarity 74.3%; Pred. No. 56;
 Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 AGGATCAACAAATGCTTACGCTTTTACCAT 36
 Db 328 AAGAGAGAACATATATAGTTCTTTAGTAT 362

RESULT 8
 AA922889 484 bp mRNA EST 19-MAY-1998
 LOCUS
 DEFINITION o151g12.s1 NCI-CGAP_HN3 Homo sapiens cDNA clone IMAGE:1486246 3' similar to SW:EC01_HUMAN O13011 PROBABLE PROXIMAL ENOYL-COA HYDROLASE ; mRNA sequence.

ACCESSION AA922889
 NID 93070198
 VERSION AA922889.1 GI:3070198
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI/NCBI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute / National Institute of Dental Research, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 On Oct 20, 1997 this sequence version replaced gi:2520646.

JOURNAL COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: John Enslay, M.D.
 CDNA Library Preparation: Stratagene, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILM at: www.bio.lnl.gov/btrp/image/image.html

Insert Length: 1258 Std Error: 0.00
 Seq primer: -40m13 fwd. Et from Amersham
 High quality sequence stop: 260.

FEATURES
 source
 1 - 484
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:1486246"
 /clone_lib="NCI-CGAP_HN3"
 /tissue_type="squamous cell carcinoma from base of tongue"
 /lab_host="SOBR (kanamycin resistant)"
 /note="Organ: tongue; Vector: Bluescript SK-; Site: 1; EcorI: Site 2: XhoI; Cloned unidirectionally. Primer: 014go dt. Average insert size 1.0 kb. 5' adaptor sequence: 5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5' (GA)10ACTACTCGAGCTTTTCTTTTCTTTT 3' "

BASE COUNT 99 a 134 c 137 g 114 t
 ORIGIN

Query Match 54.2%; Score 20.6; DB 40; Length 484;
 Best Local Similarity 74.3%; Pred. No. 57;
 Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 GATCAGAACAAATGCTTACGCTTTTACCATCA 38
 Db 446 GACCAGCCCAATGCCAGGCGCTCTGACGACATCA 480

RESULT 9
 HSM010166/c standard; RNA; EST; 700 BP.

AC AL045316;
 AC AL045316.1
 SV e1403790

DI 12-MAR-1999 (Rel. 59, Created)
 DI 12-MAR-1999 (Rel. 59, Last updated, Version 1)

DE Homo sapiens mRNA; EST DKEP434A015_r1 (from clone DKEP434A015)

XX EST; expressed sequence tag.

OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

[1]

RA Dueterhoeft A., Lauber J., Mewes W., Gassenhuber J., Wiemann S.;
 Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
 RL MRP5, Am Klopferplatz 18a D-82152 Martinsried, GERMANY

CC Clone from S. Wiemann, sequenced by Qiagen within the CDNA sequencing consortium of the German Genome Project
 CC s1 sequence also available
 CC This clone is available at the R2PD in Berlin
 CC Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

XX Key Location/Qualifiers
 FH 1. 700
 FT source
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT /clone_image="IMAGE:1486246"
 FT /clone_lib="NCI-CGAP_HN3"
 FT /tissue_type="squamous cell carcinoma from base of tongue"
 FT /lab_host="SOBR (kanamycin resistant)"
 FT /note="Organ: tongue; Vector: Bluescript SK-; Site: 1; EcorI: Site 2: XhoI; Cloned unidirectionally. Primer: 014go dt. Average insert size 1.0 kb. 5' adaptor sequence: 5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5' (GA)10ACTACTCGAGCTTTTCTTTTCTTTT 3' "

SO Sequence 700 BP; 164 A; 156 C; 125 G; 254 T; 1 other;

Query Match 53.7%; Score 20.4; DB 54; Length 700;
 Best Local Similarity 80.0%; Pred. No. 72;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGATCAACAAATGCTTACGCTTTTCTTT 30
 Db 523 AAGATCAACAAATGCTTACGCTTTTCTTT 494

RESULT 10

LOCUS R81085 717 bp mRNA EST 12-JUN-1996
 DEFINITION EST222601 WATM1 Homo sapiens cDNA clone 22201, mRNA sequence.

ACCESSION R81085
 NID 9857688
 VERSION R81085.1 GI:857688

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 717)
 AUTHORS Bouillaud, F.
 TITLE Study of expressed sequences tags in adipose tissue 1995
 JOURNAL Unpublished (1995)
 COMMENT On May 8, 1995 this sequence version replaced gi:801224.

Contact: Frederic Bouillaud
 Centre de Recherche sur l'Endocrinologie moleculaire et le
 developpement
 CNRS
 5, Rue Jules Hetzel, Meudon Bellevue, 92190 France
 Tel: 33 1 45 07 52 87
 Fax: 33 1 45 07 58 90
 Email: bouillaud@infobio.fr
 automatic cycle sequencing of PCR amplified insert
 Insert length: 851 Std Error: 0.00
 High quality sequence stop: 290.

FEATURES
 source
 1. 717
 /organism="Homo sapiens"
 /strain="caucasian"
 /db_xref="taxon:9606"
 /clone="22e01"
 /note="Vector: lambda gtl1; Site 1: EcoRI; White adipose
 tissue, sub cutaneous, adult, female. Purification of
 polyA mRNA, first strand priming with random
 oligonucleotides. Ligation of adaptors EcoRI/NotI, cloning
 in EcoRI site of lambda gtl1."

BASE COUNT 196 a 172 c 192 g 157 t
 ORIGIN

Query Match 53.2%; Score 20.2; DB 22; Length 717;
 Best Local Similarity 88.0%; Pred. No. 87;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 14 ATGCTAGCGCTTCTTACCATCA 38
 ||||| ||||| ||||| ||||| ||
 Db 258 ATGCCAGCGTTTCATTAGCCACCA 234

RESULT 11
 H84400 426 bp mRNA EST 13-NOV-1995
 LOCUS Y101a10.r1 Soares retina N2b5HR Homo sapiens cDNA clone
 DEFINITION IMAGE:223002 5' similar to gb:D90209 DNA-BINDING PROTEIN TAKREB67
 (HUMAN); mRNA sequence.
 ACCESSION H84400
 NID 91063071
 VERSION H84400.1 GI:1063071
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 426)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasaks, E., Waterston, R., Williamson, A., Wohlmann, P., and
 Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT On May 18, 1995 this sequence version replaced gi:811191.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800

Fax: 314 286 1810
 Email: est@wustl.edu
 High quality sequence stops: 260
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1659 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 260.

FEATURES
 source
 1. 426
 /organism="Homo sapiens"
 /db_xref="GDB:3854036"
 /db_xref="taxon:9606"
 /clone="IMAGE:223002"
 /clone_1ib="Soares retina N2b5HR"
 /sex="male"
 /tissue_type="retina"
 /dev_stage="55 year old"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: eye; Vector: pTZ19 (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTCACATCGAGTGGAGCGCGCCGCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pTZ19 vector
 (Pharmacia). The retinas were obtained from a 55 year old
 caucasian and total cellular poly(A)+ RNA was extracted 6
 hrs after their removal. The retina RNA was kindly
 provided by Roderick R. McInnes M.D. Ph.D. from the
 University of Toronto. Library constructed by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 89 a 117 c 113 g 99 t 8 others
 ORIGIN

Query Match 53.2%; Score 20.2; DB 24; Length 426;
 Best Local Similarity 73.5%; Pred. No. 83;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 AAGGATCAGACATGCGCTTCTTACCC 34
 ||||| ||||| ||||| ||||| ||
 Db 371 AATCATCTTAGAGACCGCTTCTTACCC 338

RESULT 12
 AA018084 400 bp mRNA EST 30-JAN-1997
 LOCUS zef1a10.s1 Soares retina N2b4HR Homo sapiens cDNA clone
 DEFINITION IMAGE:362490 3' mRNA sequence.
 ACCESSION AA018084
 NID 91481339
 VERSION AA018084.1 GI:1481339
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 400)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Scheilenberg, K., Soares, M.B., Tan, F., Tillery-Meg, J.,
 Trevasaks, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
 and Matra, M.
 TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478
 COMMENT On Apr 14, 1993 this sequence version replaced gi:785894.

Contact: Wilson RK


```

source
1.307
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="423327"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15'
TCTTACCAATCTGAAAGTGGAGCGCGGGAATTTTCTTTTCTTTT
T 3', on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bernaldo."

BASE COUNT      111 a      49 c      46 g      101 t
ORIGIN

Query Match      52.6%; Score 20; DB 46; Length 307;
Best Local Similarity 82.1%; Pred. No. 98;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy      3 GGATCAGACATGCGCTAGGCTTCTTT 30
Db      159 GAATCAGAAAATGCCAGCCTTTT 132

RESULT 15
AA799278      628 bp      mRNA      EST      30-APR-1998
LOCUS      AA799278/c
DEFINITION      EST188775 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
ACCESSION      RHEA05 5' end, mRNA sequence.
ID      AA799278
VERSION      92862233
KEYWORDS      AA799278.1 GI:2862233
SOURCE      EST
ORGANISM      Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
Lee,N.H., Glodet,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
Unpublished (1998)
On Jan 19, 1998 this sequence version replaced gi:2286418.
Other ESTs: EST188774
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13 Reverse.
Location/Qualifiers
1..628
/organism="Rattus sp."
/db_xref="ATCC (1host):2006305"
/db_xref="taxon:10118"
/clone_lib="RHEA05"
/clone_image="Normalized rat heart, Bento Soares"
/notes="Organ: heart; Vector: pT73pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT      169 a      163 c      155 g      140 t      1 others
ORIGIN

Query Match      52.1%; Score 19.8; DB 38; Length 628;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy      1 AAGATCAGACATGCGCTAGGCTTCTTTA 31
Db      493 AGGATTAGAACAAATCCCTCCTTCTTTA 463

Search completed: September 13, 1999, 16:14:01
Job time: 5460 sec

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:59 ; Search time 539.84 Seconds

(without alignments)
135,498 Million cell updates/sec

Title: US-09-325-095-29

Sequence: 1 RAARTGCCRAARTCDATRTGMA 23

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sy:*
15: gb_un:*
16: gb_un:*
17: gb_y1:*
18: em_fun:*
19: em_hlg:*
20: em_hum1:*
21: em_hum2:*
22: em_hn:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_st:*
31: em_sy:*
32: em_un:*
33: em_v1:*
34: gb_hlg1:*
35: gb_hlg2:*
36: gb_hn1:*
37: gb_hn2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	ID	Description
1	18.2	79.1	111597	11 HS085196
				US8196 Homo sapien

C	2	18.2	79.1	247877	11	HUAE000661	AE000661 Homo sapi
C	3	18.2	79.1	42467	37	CELF26F12	U55373 Caenorhabdl
C	4	18.2	79.1	37675	37	CELM01G5	AF078786 Caenorhab
C	5	17.4	75.7	297070	34	CEY47H10	Z95311 Caenorhabdl
C	6	17.4	75.7	21816	36	CECA47F8	AL009246 Caenorhab
C	7	17.2	74.8	155147	9	HS992D9	AL008638 Human DNA
C	8	17.2	74.8	218336	11	AF067844	AF067844 Homo sapi
C	9	17	73.9	97015	1	BSEGENR	X73124 B. subtilis
C	10	17	73.9	212150	1	BSUB0020	Z99123 Bacillus su
C	11	17	73.9	3650	7	YSCEMET10A	L26504 Saccharomyc
C	12	17	73.9	126839	10	HS19408	AL031769 Human DNA
C	13	17	73.9	7215	36	DMP13K68D	X92892 D. melanogas
C	14	17	73.9	10213	36	PFU27338	U27338 Plasmodium
C	15	17	73.9	43584	37	CELB0207	U97196 Caenorhabdl
C	16	17	73.9	3606	37	DMU52192	U52192 Drosophila
C	17	16.6	72.2	3606	4	AB008857	AB008857 Fugu rubr
C	18	16.6	72.2	38	5	A37242	A37242 Sequence 11
C	19	16.6	72.2	87841	7	AB005247	AB005247 Arabidops
C	20	16.6	72.2	85791	7	AB016888	AB016888 Arabidops
C	21	16.6	72.2	69748	7	SCRA011	X78993 S. cerevisia
C	22	16.6	72.2	1600	7	SCYBR107C	Z35976 S. cerevisia
C	23	16.6	72.2	3390	7	SCYBR108W	Z35977 S. cerevisia
C	24	16.6	72.2	2224	7	SCYOL033W	Z74775 S. cerevisia
C	25	16.6	72.2	25380	7	SPAC1P8	Z81312 S. pombe chr
C	26	16.6	72.2	2043	7	YSCME51G	L39015 Saccharomyc
C	27	16.6	72.2	3485	8	AF025434	AF025434 Papaver s
C	28	16.6	72.2	126337	8	AP000067	AP000067 Arabidops
C	29	16.6	72.2	85444	9	AP000023	AP000023 Homo sapi
C	30	16.6	72.2	153147	9	HS13D10	AL021407 Homo sapi
C	31	16.6	72.2	126474	9	HS181N1	Z82899 Human DNA s
C	32	16.6	72.2	130705	9	HS232L22	Z73986 Human DNA s
C	33	16.6	72.2	113704	9	HS267P19	Z75889 Human DNA s
C	34	16.6	72.2	132229	9	HS447N6	Z82211 Human DNA s
C	35	16.6	72.2	31523	9	HS656F14	AL034416 Human DNA
C	36	16.6	72.2	123004	10	HS196E23	Z97632 Human DNA s
C	37	16.6	72.2	95855	10	HSAC000115	AC000115 Human BAC
C	38	16.6	72.2	39876	10	HS083C4	Z70050 Human DNA s
C	39	16.6	72.2	41052	11	AC002499	AC002499 Human Cos
C	40	16.6	72.2	41876	11	AC004190	AC004190 Homo sapi
C	41	16.6	72.2	746	12	AF084643	AF084643 Mus muscu
C	42	16.6	72.2	1460	17	CAEYTRNA	X64109 Caprine Art
C	43	16.6	72.2	1999	17	OLYENYORK	L19198 Ovine lenti
C	44	16.6	72.2	144247	34	AC005482	AC005482 Homo sapi
C	45	16.6	72.2	144247	34	AC005482	AC005482 Homo sapi

ALIGNMENTS

RESULT	1	HS085196	111597 bp	DNA		
LOCUS		HS085196	Homo sapiens BAC378, complete sequence.		PRI	27-MAY-1997
DEFINITION		US8196				
ACCESSION		US8196.1	GI:2121230			
NID		92121230				
VERSION		US8196.1	GI:2121230			
KEYWORDS		HTG.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS		Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
TITLE		Boysen, C., Simon, M.I., and Hood, L.				
JOURNAL		Analysis of the 1.1-Mb human alpha-T-cell receptor locus with				
MEDLINE		bacterial artificial chromosome clones				
REFERENCE		Genome Res. 7 (4), 330-338 (1997)				
AUTHORS		2 (bases 1 to 111597)				
TITLE		Boysen, C., Simon, M.I., and Hood, L.				
JOURNAL		Submitted (14-JAN-1997) Molecular Biotechnology, University of				
FEATURES		Washington, Box 357730, Seattle, WA 98195, USA				
		Location/Qualifiers				

```

source          1. 111597
                 /db_xref="taxon:9606"
source          44957. 46288
                 /db_xref="taxon:9606"
                 /insertion_seq="152"
BASE COUNT     33734 a 23173 c 21631 g 33059 t
ORIGIN
Query Match    79.1%; Score 18.2; DB 11; Length 111597;
Best Local Similarity 65.2%; Pred. No. 1e+02;
Matches 15; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Oy 1 RAARTGCCCAARTCDARTGRGA 23
Db 47404 AAAAACTAAATCAATATGAAA 47382
      ||| ||| ||| ||| ||| |||
RESULT 2
LOCUS       HUAE000661/c          DNA          03-JUN-1998
DEFINITION  Homo sapiens T-cell receptor alpha delta locus from bases 752679 to
ACCESSION   AE000661 AE000521
VERSION     92358060
KEYWORDS    AE000661.1 GI:2358060
SOURCE      human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 247877)
AUTHORS     Koop,B.F., Rowen,L., Wang,K., Kuo,C.L., Seto,D., Lenstra,J.A.,
            Howard,S., Shan,W., Deshpande,P., and Hood,L.
            The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region:
            organization, sequence, and evolution of 97.6 kb of DNA
            Genomics 19 (3), 478-493 (1994)
            94245236
            This citation covers from bases 966383-1064019
            2 (bases 1 to 247877)
            Boysen,C., Simon,M.I., and Hood,L.
            Analysis of the 1.1-kb human alpha/delta T-cell receptor locus with
            bacterial artificial chromosome clones
            Genome Res. 3(30), 330-338 (1997)
            3 (bases 1 to 247877)
            Boysen,C., Inyoub,L., Smith,T.M., Smit,A., Wang,K., Rowen,L. and
            Hood,L.
            T-Cell Receptor Alpha Delta Locus Complete Nucleotide Sequence
            Unpublished (1997)
            This citation covers bases 1-983545 and bases 1064020-1071650
            4 (bases 1 to 247877)
            Boysen,C., Lee,I., Smith,T.M., Smit,A., Kai,W., Lee,R. and Leroy,H.
            Direct Subcloning
            Submitted (20-JUN-1997) Department of Molecular Biotechnology,
            University of Washington, Box 357730, Seattle, Washington 98195,
            USA
REMARK       Complete nucleotide sequence of the human T-cell receptor alpha
            delta locus
COMMENT      *Sequencing method: high redundancy shotgun. Interspersed Repeats
            were identified with RepeatMasker (available from
            http://ftp.genome.washington.edu/RV/RepeatMasker.html) Simple
            sequence repeats were identified with sputnik (available from
            http://serc.mbt.washington.edu/~christa/software/sputnik.html).
            Location/Qualifiers
FEATURES             1. 247877
source              /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   <1. 92762
                   /note="This span of sequence was submitted independently
                   as Genbank Accession Number U85196."
                   /organism="Homo sapiens"

```

```

repeat_region    /db_xref="taxon:9606"
repeat_region    /clone="BAC 378"
repeat_region    /clone_lib="Mel Simon's BAC library"
repeat_region    complement(1042..1338)
repeat_region    /rpt_family="AluB"
repeat_region    complement(1717..2016)
repeat_region    /rpt_family="AluSx"
repeat_region    complement(2061..3002)
repeat_region    /rpt_family="L1MC1"
repeat_region    complement(3677..3796)
repeat_region    /rpt_family="MIR"
repeat_region    complement(4196..4512)
repeat_region    /rpt_family="AluSg"
repeat_region    5116..5415
repeat_region    /rpt_family="AluSg"
repeat_region    5405..5410
repeat_region    /rpt_type="tandem"
repeat_region    /rpt_unit="AAG"
repeat_region    5994..6165
repeat_region    /rpt_family="AluB"
repeat_region    complement(9217..10108)
repeat_region    /rpt_family="L1PA2"
repeat_region    complement(9956..13678)
repeat_region    /rpt_family="L1"
repeat_region    complement(13678..14564)
repeat_region    /rpt_family="L1"
repeat_region    complement(14678..15378)
repeat_region    /rpt_family="MIR2"
repeat_region    complement(16264..16634)
repeat_region    /rpt_family="L1"
repeat_region    18338..19151
repeat_region    /gene="TCRAV37S1"
repeat_region    /pseudo
repeat_region    18338..18357
repeat_region    /gene="TCRAV37S1"
repeat_region    /standard_name="TCRAV37S1"
repeat_region    /pseudo
repeat_region    join(18598..18649,18821..>19113)
repeat_region    /gene="TCRAV37S1"
repeat_region    /standard_name="TCRAV37S1"
repeat_region    /pseudo
repeat_region    /product="TCRAV37S1"
repeat_region    join(18598..18649,18821..>19113)
repeat_region    /gene="TCRAV37S1"
repeat_region    /standard_name="TCRAV37S1"
repeat_region    /pseudo
repeat_region    19114..19120
repeat_region    /gene="TCRAV37S1"
repeat_region    /standard_name="TCRAV37S1"
repeat_region    /note="heptamer"
repeat_region    19121..19142
repeat_region    /gene="TCRAV37S1"
repeat_region    /standard_name="TCRAV37S1"
repeat_region    /note="spacer"
repeat_region    19143..19151
repeat_region    /gene="TCRAV37S1"
repeat_region    /standard_name="TCRAV37S1"
repeat_region    complement(21888..22551)
repeat_region    /note="nonamer"
repeat_region    /rpt_family="MIR50"
repeat_region    complement(22752..23305)
repeat_region    /rpt_family="MIR4B"
repeat_region    complement(23900..24202)
repeat_region    /rpt_family="AluSg"
repeat_region    join(24820..24868,25112..>25412)
repeat_region    /gene="TCRAV38S1"
repeat_region    /standard_name="TCRAV38S1"
repeat_region    /note="Old name TCRAV38S1"
repeat_region    /product="TCRAV38S1"
repeat_region    join(24820..24868,25112..>25412)
repeat_region    /gene="TCRAV38S1"
repeat_region    /standard_name="TCRAV38S1"
repeat_region    /note="Old name TCRAV14S2"

```

```

gene
24820..25451
/gene="TCRAV38S1"
join(24820..24866,25112..>25412)
/gene="TCRAV38S1"
/standard_name="TCRAV38S1"
/note="Old name TCRAV14S2"
/codon_start=1
/product="TCRAV38S1"
/protein_id="AAB69035.1"
/db_xref="PIR:G2358061"
/db_xref="GI:2358061"
/translation="MTRVSLMAVVSTCLESGMAQVYTOGPEKSVQEAETVTLCT
YDTSNNYILEFVKQPPSRQMLIVIRQNAVYQQNATNFRFSVNFQKAAKSLIKISDS
QUGDPAMTFCAPMK"
misc_recomb
25413..25419
/gene="TCRAV38S1"
/standard_name="TCRAV38S1"
/note="heptamer"
25420..25442
/gene="TCRAV38S1"
/standard_name="TCRAV38S1"
/note="spacer"
25443..25451
/gene="TCRAV38S1"
/standard_name="TCRAV38S1"
/note="nonamer"
complement(27096..27964)
/rpt_family="LIPB3"
complement(27821..29134)
/rpt_family="L1"
complement(30712..31212)
/rpt_family="KLT2CA"
32458..32606
/rpt_family="AluJb"
33599..33609
/rpt_type=tandem
/rpt_unit=86
join(<34000..34048,34298..>34597)
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="Old name TCRAV14S1-hDV8S1"
/product="hADV38S2"
34000..34636
/gene="hADV38S2"
join(34000..34048,34298..>34597)
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="Old name TCRAV14S1-hDV8S1"
/codon_start=1
/product="hADV38S2"
/protein_id="AAB69036.1"
/db_xref="PIR:G2358062"
/db_xref="GI:2358062"
/translation="MACPGFLMALVYISICLESNAQVYTOGPEKSVQEAETVTLCT
YDTSDDYLEYKQPPSRQMLIVIRQNAVYQQNATNFRFSVNFQKAAKSPSLIKISDS
QUGDAMATFCAPKS"
V_segment
join(34000..34048,34298..>34597)
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="Old name TCRAV14S1-hDV8S1"
34598..34604
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="heptamer"
34605..34627
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="spacer"
34628..34636
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="nonamer"
36263..36614
repeat_region

```

```

repeat_region      /rpt_family="THEIA"
                   complement(36621..36703)
                   /rpt_family="MIR2"

Query Match      79.1% Score 18.2; DB 11; Length 24787;
Best Local Similarity 65.2% Pred. No. 1.1e+02;
Matches 15; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 BAARTGCCRAARTCDATRTGAA 23
      :|||:|:|:|:|:|:|:|:|:|
Db 28569 AAAAGCTAAATCAATATGAA 28547

RESULT 3
CELF26F12/c 42467 bp DNA INV 02-JUN-1999
LOCUS
DEFINITION Caenorhabditis elegans cosmid F26F12.
ACCESSION U55373
NID g3294506
VERSION U55373.1 GI:3294506
KEYWORDS
SOURCE
ORGANISM
.
Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 42467)
The C. elegans Genome Sequencing Consortium, Washington University
Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
Hinxton, U.K.,C.

TITLE Genome sequence of the nematode C. elegans: a platform for
JOURNAL Investigating biology. The C. elegans Sequencing Consortium
MEDLINE Science 282 (5396), 2012-2018 (1998)
REMARK Erratum: [[published erratum appears in Science 1999 Jan
1:283(5398):351]]
2 (bases 1 to 42467)
Wilson, R., Bentley, D. and Gattung, S.
The sequence of C. elegans cosmid F26F12
JOURNAL Unpublished (1999)
REFERENCE 3 (bases 1 to 42467)
Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (17-APR-1996) Robert Waterston
REFERENCE 4 (bases 1 to 42467)
Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 42467)
Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 8, 1998 this sequence version replaced gi.1280109.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1HQ, England
e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (1.e., phred quality >=
30); an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by

```

sequence from more than one m13 subclone.

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is H03613, 1800 bp overlap; 3' cosmid is C0266, 200 bp overlap. Actual start of this cosmid is at base position 1601 of C0266f12; actual end is at 42467 of C0266f12

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation). Location/Qualifiers

FEATURES

source

1. 42467

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/clone="F26f12"

/chromosome="V"

74. 5907

/gene="F26f12.7"

join(74. 403.473. 1465.1514. 2592.2644. 3973.4045. 5390. 5442. 5713.5768. 5907)

CDS

gene

/note="contains similarity to Pfam domains: PF00628 (PHD),

Score=75.0, E-value=5e-19, N=2; PF00176 (SNF_N),

Score=343.1, E-value=9.8e-100, N=1; PF00385 (chromo),

Score=26.8, E-value=7.6e-07, N=2; PF00271 (helicase_C),

Score=93.9, E-value=1e-24, N=1; coded for by C. elegans

cDNA yk283.5; coded for by C. elegans cDNA yk18012.5;

coded for by C. elegans cDNA yk21a12.3; coded for by C.

elegans cDNA yk21a12.5; coded for by C. elegans cDNA

yk28d2.3; coded for by C. elegans cDNA yk28d2.5; coded for

by C. elegans cDNA yk28h7.3; coded for by C. elegans cDNA

yk28h7.5; coded for by C. elegans cDNA yk34h7.3; coded for

by C. elegans cDNA yk34h7.5; coded for by C. elegans cDNA

yk49c2.3; coded for by C. elegans cDNA yk49c2.5; coded for

by C. elegans cDNA yk72d6.3; coded for by C. elegans cDNA

yk72d6.5; coded for by C. elegans cDNA yk131b5.3; coded

for by C. elegans cDNA yk131b5.5; coded for by C. elegans

cDNA yk148d9.3; coded for by C. elegans cDNA yk172f1.3;

coded for by C. elegans cDNA yk172f1.5; coded for by C.

elegans cDNA yk177h5.3; coded for by C. elegans cDNA

yk177h5.5"

/codon_start=-1

/protein_id="AAC25894.1"

/db_xref="PID:g3294508"

/db_xref="GI:3294508"

/translation="MSTEDPSLVDAESMSEGSVYQATETETEEDEOQDEAGPS

ERRSSRRKGGKGRKSAASAKVEIPDPYSTSEVCAAGLTVDEDDYDEE

FOGILNLTFSIIRPOLLEAPNTVNSKMPQVKEYEDHAAOQKPOKQAR

STIPAVSTPIIPPSAPPTTSGARRRRSDAPSDOEBAFIRKQOEEDLVNKE

DARIRAREERKKKGALEAARAKKATKEGEAENDYCECKODSGELITCDPCR

AVTICIDENMEPEEGDMSCHACTEHPVEYKEPAKONDFCTICKETETLLCDS

CVCSTHATCIDPPLTVEPKETWSPCEYVAPKEIKETILCWKKEIIPPEPLAEG

EASDDAMLPKPRKMEPREREFVWKWKLISWQSVSEMLLEHFMALILYWRKN

DSADPEEESTVSRSHSDNDYKLRERYOIGIRPEWQIIRHINHO SYKASQDYL

VYMKELSDOATMERDSDNIANYEALIKYHGRSKLMEDIKRWAKIAKRRAGK

LPPKDEKKKKRERKIDIRKRYEOPDYVTEGGLHPQLEGMLWLRHGWCKRAG

LADMGIGKTVOSLITFLYSIMKEGCKGPPFLAALSTIINNERAEDWCPDYVYV

IGLRARVYLREHRESFVEGAVSGPKASKMTTNMKEFVLLTSETINDKTLSS

IEWGLVYDEAHLRKNOSLFKKNLEYIIRHVLITGPIONNLEELFHLNLTSSK

RFNOLEATAEENEISKEDQIEKLNHLGPHLRRLKADVLGMPKSELIVRYELSA

MOKWKYKILITENFDALVNGSGTQMSLMNYLMELKCCNHPYLKYAKELAPRKNG

MYEGTALIKSGKFTYLOKMLKLEKIDGGHRLVLSOMRMLDIMEGLCYBGRYERI

DGSIQGRKODIDIKYNAFGROQIFLSTRAGGGLINLATADYITITDSQNPNDI

QAFSAHLRIGQHKWYIRFVYTKRSEKRTISVAKKMLNLVVRAGLGGEGKWTMS

KTELDVLRWGTSELSDLDAAEGESKKAAQOETVMDAAVADLALDRSKKETP

AGEDEERAEQNEEYLSFFKVASYOTKEEGDEEEDEEYIKRDEKEDPDYKEL

LKHHEODREIELKLGKGRYRKQINYASEMGWDMSONOTODDDNDSYRSDN

DGGLNSDEDDYDEKKRRRDEKMPPLAKVNGVYELLGFNRQKATYGAVMRNGMP

gene

CDS

/gene="F26f12.6"

join(7149. 7372.7431. 7576.7626. 8132.8178. 8350)

/note="similar to C. elegans ODR-10-related proteins"

/codon_start=-1

/evidence="not-experimental"

/protein_id="AAC25893.1"

/db_xref="PID:g3294507"

/db_xref="GI:3294507"

/translation="MSAGSLDEYFRKNYTVGCEKSTSEKSTDFIYPLQSTLNT

PLEVLGYCIIQKTPKRSIGSTKMAVFNHIMCQSLSTLPLPIWLPYIGSSMG

LFVYIGVEYVQYLFVYVMEFTWSSVSPENHNALIQHTLIRISTSVRYLSAN

LYVSEFPFPLFNLINPQESAKLILIQVPCPOEEFPLSVFISTDANFRLYVSM

LIFLCSTVQOLLFYAFCTLRVLRSSSPSEKRTIELQKFFIGLILQVNPVAFYV

SVYSCISVYKNTYNNQMLNPFVLYVNFYGVSTIGLISCHTAIRWKFARKLMKRSRT

DLASNTSVSRNNQID"

complement(9107. 10024)

/gene="F26f12.1"

complement(join(9107. 9841.9890. 10024))

/note="similar to cuticular collagen; coded for by C.

elegans cDNA yk92h9.3; coded for by C. elegans cDNA

yk100f8.5; coded for by C. elegans cDNA yk123h6.5; coded

for by C. elegans cDNA yk125b5.5; coded for by C. elegans

cDNA yk125b5.3"

/codon_start=-1

/protein_id="AAC25888.1"

/db_xref="PID:g1280114"

/db_xref="GI:1280114"

/translation="MGKILIVGSSIALTCLVYFTMGSLQIEISDKMKEVEDGMER

REIHODIYNVRRKHINPTGETDAPNPQTLFGRRSOSFPPCCNGCPGSECPG

PPGPGPBGESGPDGDKPLGALVYVAVIYDLPGGCICPGPRGPGPGPGV

PAGSGGNGRGKPGPGPGPGPGPGSGDGGGRGPRGPRGPRGPGRETETKPPQGR

PGPPGPRGAPPAQPGSPGMDGSGKRGNAGRGPPGAPKNGTPGVRGDAAPGD

AGVCPARAVAK"

complement(10850. 11800)

/gene="F26f12.2"

complement(join(10850. 10922.10979. 11185.11236. 11520.

11566. 11705.11753. 11800))

/gene="F26f12.2"

/codon_start=-1

/evidence="not-experimental"

/protein_id="AAC25889.1"

/db_xref="PID:g1280113"

/db_xref="GI:1280113"

/translation="MTEFVCGDNQSAEQKNPNQNPFFKYAPNKKQALPILLODY

CINPISLRYCYAPNOSLITLWMDKRISEYFEKALILEKLIQNNNLARPK

LNNIDAPNNEKSPISQSLNVAVPTKMLNLEKYLAKYKAVILTAAXQYGVFYH

CKVOSKRSKSLINISKMLFNNQDEADYLKGTGLYALHLPDERKMYKIEREYK

LKLNIRKIVADILVLDIGGYVSGRTVFE"

complement(113045. 17228)

/gene="F26f12.3b"

/gene="F26f12.3b"

/complement(113045. 18888)

/gene="F26f12.3"

/complement(113045. 18888)

/gene="F26f12.3a"

/complement(join(113045. 13126.13675. 13808.13860. 15943.

17111. 17445.18851. 18888))

/gene="F26f12.3"

Query Match 79.1%; Score 18.2; DB 37; Length 42467; Best Local Similarity 65.2%; Pred. No. 97; Matches 15; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY 1 RAATGCCAARTCDATRTGRA 23
 DB 29885 AAAATGCCAAAATGATAGGAAA 29863

RESULT 4

LOCUS CELM01G5 37675 bp DNA INV 21-JUL-1998
 DEFINITION Caenorhabditis elegans cosmid M01G5.
 ACCESSION AF078786
 MID G3329634
 VERSION AF078786.1 GI:3329634
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE 1 (bases 1 to 37675)
 AUTHORS Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershav, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Ritten, L., Roopra, A., Saunders, D., Showkseen, R., Smalton, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstein, L., Wilkinson-Sprout, J. and Mohlman, P.
 TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

JOURNAL Nature 368 (6466), 32-38 (1994)
 MEDLINE 94150718
 REFERENCE 2 (bases 1 to 37675)
 AUTHORS Geisel, C., Kramer, J. and Tyman, B.
 TITLE The sequence of C. elegans cosmid M01G5
 JOURNAL Unpublished (1998)
 REFERENCE 3 (bases 1 to 37675)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUL-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 CUMULANT Submitted by:

Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1HQ, England
 e-mail: twenematode.wustl.edu and jens@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternative chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is Y7ID11. Actual start of this cosmid is at base position 1 of CELM01G5; actual end is at 37675 of CELM01G5.

NOTES:

Coding sequences below are predicted from computer analysis, using

FEATURES

the program GeneFinder.P. Green and L. Hillier, ms in preparation).
 Location/Qualifiers

1..37675

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/clone="M01G5"

/chromosome="III"

2941..13464

/gene="M01G5.1" .3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,11320..13464)

/gene="M01G5.1"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC2694.1"

/db_xref="PID:g3329635"

/db_xref="GI:3329635"

/translation="MRKCPHFHFPPKNDLFFPKNVNPFPEPSEKSDQARKYOK
 ROYHQMTPEEFDFLIESDFGDMCHETAKFRIRVAMVLRRLRKQREERLE
 AEKTAIRIKREDEDEKREAEKLMDEKRLIOEKQKDEKQDEKIDKHELEKRVLE
 AAEKREPSLEPGLVVEPRRHOKVYKNSMPAPSPYVLCEHSTIEPHQCKM
 NAMSNLMYGERQHTSNRELRKIDRATEREYDVLVALEHKNACGHPDVDR
 KIOKCLDKRNRPPEPCENLOFPEKDPVYIYENVDADGIRGINOALVPTKILRG
 VMNVLLPDISQKPLNLEDSTPPPKRLDLDALASITIAENPTVIMADATTIC
 HLTPNNESSDYDTIPCVAREENLPEVGNKVVLSLPPMSGITIASIQNSLKA
 LRSQFLKSKPKIEKPVVDAIYEVTIESEAPAEAGATDFDDLTTEATKSKK
 EAKNEKIPENORQSTVOYGFIRGDAIRLVSNAPYGVDEPKKNAFLPHSSORO
 AOAMVTFEPRRIEPLPMGAMDMGAPGNMTRKAVLKMDSHILYTSYRLHILDI
 DALTKAVDVEPPPGAGFLNCKSTISIFPEPPIEKRRKVKLTFFQYVLSACVHL
 EKLINLETNLFGEYIVHFNKPLVITPOCEKGVITEARLTGNVOSSDPAAGS
 SFYGFSPDLPEQWQIITQGRAPRLAKDSPLANLTKQNFNIQWKSQVKKRYED
 YSAREDEKRNRLTDQIDPNISADFTNPAMI PVFEKRPFPVRVGRGGRGSGGSG
 GGRGRERDGSRGGRGSGRGNSEBPVADIV"

complement(14420..17884)

/gene="M01G5.3"

complement(14420..14553,15666..15883,17806..17884)

/gene="M01G5.3"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC2694.1"

/db_xref="PID:g3329637"

/db_xref="GI:3329637"

/translation="MVVVTSLPVVCCAMCAHLERWVLSGISLPLHLHLARV
 GSKSVSLRPPPTFPPTGTSNGYCGIYVAIORINIFVAFPGWICIDSPVABA
 HRRKRYHHPGGLXSEOVNHFVYKHYHIDPKYQ"

complement(19106..24912)

/gene="M01G5.4"

complement(19106..19411,20485..20540,20660..20829,
 22666..22784,22856..22918,24805..24912)

/gene="M01G5.4"

/note="coded for by C. elegans cDNA yk479a10.5; coded for
 by C. elegans cDNA yk237e8.5"

/codon_start=1

/protein_id="AAC26942.1"

/db_xref="PID:g3329636"

/db_xref="GI:3329636"

/translation="MEKNLEKPKKKIFFPGIFPSKNOYFEPSPNNMPATKELDS
 EKKVODPOKLEKIDHEVYKNSROVAKPOETGAKKHNNYLLKNKEMROKQSA
 EKLEDAETSKTDVLSHAKLEBELAVANLKRDMQGLPQATNFYINLIGSLANS
 LITAEKFKRKQKTEFFENLFEHLSFEFFQYVHFLPLTKTAIFLDLYRSEVAY
 FRRYSNPENIKTSDPELVTKPGKLNKYNNSNFMENMLNFFPEKIDKKNL"

complement(26448..336757)

/gene="M01G5.5"

complement(10126448..26861,28077..28620,30099..30337,
 31338..31470,33258..33370,34744..34919,36697..36757)

/gene="M01G5.5"

/note="similar to the sodium:neurotransmitter symporter
 family (Pfam: SMF_hmm, score: 34.33); most similar to
 sodium- and chloride-dependent glycine transporters"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC26944.1"

```

/db_xref="PID:93329638"
/db_xref="GI:3329638"
/translation="GWSNKRDIYLVIGTACGVSFWKRPPIVPOHGAAFLVPIYCL
MCLASLPHFPKENTLQGFSSASDYSVWRKVPLEKRGIFRAVTTISGFVAFFNITSA
TLFLINSEFSIPCNVDMLTLKSGVADTGTLMWGLGCLVCLAMIAVEICLFQGVKSS
KVVAIVAVPIIILVLVLRLLTLDGSLAAVEYELTPKEMILMDLHWGEAAVAQAFY
VSCSGGLFTTASYSRFNNNTIKYDMLIVDIVIVSLVCGILTFESAIFCTGEPAIS
DKFHRDGFHLVFEVFLAALAGVAVAPADGFEFIMLVYHATONFVENTIVSICF
EXPERLRRRHRLVLTVALILLSIPCCSSGLEFMELLQPVLPVLPVYIALECK
AINWGYDNLMDNKNWLVGTWPCPTIKWIKLFETCPWVLAIALICFLMDMSIQTH
SYQPFYWSILTWACIASPLIIPVIGWQICAGITQGMVRLVYDDDMGPMMA
HRAEFPLQIPEARLRLPPEVETIASRGVLYQVDHMERKLRVSFTPRVY"

BASE COUNT      12631 a      6500 c      6589 g      11955 t
ORIGIN

Query Match      79.1%; Score 18.2; DB 37; Length 37675;
Best Local Similarity 65.2%; Pred. No. 96;
Matches 15; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY      1 RAARTGCCRAATCARTRTGAA 23
      :11:1111:11:1:11:11:11
Db 24083 AAGTGCACAAATTATATGAA 24105

RESULT 5
CEY47H10/c      297070 bp      DNA      HTG      29-MAY-1999
LOCUS      Caenorhabditis elegans chromosome I clone Y47H10, WORKING DRAFT
DEFINITION      SEQUENCE, in unordered pieces.
ACCESSION      295311
NID      94938508
VERSION      295311.9 GI:4938508
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
ORGANISM      Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabdittina; Rhabdittidae; Rhabdittidae; Peloderiinae; Caenorhabditis
1 (bases 1 to 297070)
McLay, K.
Direct Submission
Submitted (29-MAY-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridgeshire CB10 1HQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jesssanger.ac.uk or twenemata@wustl.edu
On Jun 1, 1999 this sequence version replaced gi:4914472.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
      source
      1..297070
      /organism="Caenorhabditis elegans"
      /db_xref="taxon:6239"
      /chromosome="I"
      /clone="Y47H10"

BASE COUNT      93108 a      55103 c      92826 t      2401 others
ORIGIN

Query Match      75.7%; Score 17.4; DB 34; Length 297070;
Best Local Similarity 65.2%; Pred. No. 2,5e+02;
Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

OY      1 RAARTGCCRAATCARTRTGAA 23
      :11:1111:11:11:11:11

```

```

DB 287208  GAAATGCCAAATCCTATGAA 287186

RESULT      6
LOCUS       CEC47F8      21816 bp      DNA
DEFINITION  Caenorhabditis elegans cosmid C47F8, complete sequence.
ACCESSION   AL009246
VERSION     93217213
KEYWORDS    GI:3217213
SOURCE      HTG.
ORGANISM    Caenorhabditis elegans.
            Caenorhabditis elegans.
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
            1 (bases 1 to 21816)
REFERENCE   1
AUTHORS     McLAY, K.
TITLE       Direct Submission
JOURNAL     Submitted (04-DEC-1997) Louis, MO 63110, USA. E-mail:
            joesanger.ac.uk or rvenematode.wustl.edu
REFERENCE   2
AUTHORS     Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
            Bonfield, J., Barton, J., Connell, M., Copesey, T., Cooper, J.,
            Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
            Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
            Johnston, L., Jones, M., Kershav, J., Kirsten, J., Lister, R.,
            Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
            O'Callaghan, M., Parsons, J., Percy, C., Riffen, L., Rooper, A.,
            Saunders, D., Showkeen, R., Smailton, N., Smith, A., Sonnenhammer, E.,
            Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
            Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
            Wilkinson-Sproat, J. and Wohlman, P.
TITLE       2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
JOURNAL     Nature 368 (6466), 32-38 (1994)
MEDLINE     94150718
COMMENT     On Jun 13, 1998 this sequence version replaced gi:2664249.
            Coding sequences below are predicted from computer analysis, using
            predictions from GeneFinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see: -
            http://webace.sanger.ac.uk/cgi-
            bin/display7db-wormaceclass-Sequence6object-C47F8
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is NOT necessarily the entire insert of
            the specified clone. It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone C47F8.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone T09E11 is at 21715 in this sequence. The
            true right end of clone F56H6 is at 105 in this sequence. The start
            of this sequence (1..105) overlaps with the end of sequence Z81533
            The end of this sequence (21715..21816) overlaps with the start of
            sequence Z81147.
            Location/Qualifiers
                1..21816
                /organism="Caenorhabditis elegans"
                /db_xref="taxon:6239"
                /chromosome="I"
                /clone="C47F8"
                complement(1312..2519)
                /gene="C47F8.4"
                complement(join(1312..1413,1468..1724,2073..2401,
                2461..2519))
                /gene="C47F8.4"

```

```

/codon_start=1
/protein_id="CA15837.1"
/db_xref="PID:e1344648"
/db_xref="PID:g3875099"
/db_xref="GI:3875099"
/db_xref="SPTREMBL:062113"
/translation="MVIIVPTKRIKTRIPDMTILANTLAHVENLHMLVVEDGYIY
EVROMLEKTNSTIYMAKTRKQIPSRGHWYRTMALIRSSSAKILKONKAGYVF
ADDDNAIDVRLFTDIIRKVNNTLVGAWGLVGGVVEAPKVVNQVTAENVWALSRR
ADVAGFAINIKLILNSDAVGTDCRREGGPECLLEMDGLKMEIDIEPFGYDATKVR
DIWVHTKTSPELEOTDQPIDSLGYEVEY"
3341..4845
/gene="C47F8.6"
/join(3341..3395,3442..3540,3595..3956,4139..4338,
4383..4509,4569..4658,4705..4845)
/gene="C47F8.6"
/codon_start=1
/protein_id="CA15839.1"
/db_xref="PID:e1344650"
/db_xref="PID:g3875101"
/db_xref="GI:3875101"
/db_xref="SPTREMBL:062115"
/translation="MPRENRSLVLLPILITFETFLCKLSEKVEEPLADEAPRONC
TKNGMGSTAPETODEGSEEMISPAKOKVNTLVPTIEETSEKEDILMTIVASRD
SVARNINRQWAKNSSEIYANGRMKSLFVGLAPADYKTKRMQDAKIXGDIIV
DMDTYEBELIKSLMFLFVSKAPQIKITKIDEDLMEFPDKLALTEQGIIDSTIV
SLYGLVTPAGDIFRDKTNVYVESAVSCSQPAYLSGMYMATREAAOMLSTKR
RDFLOVEDVILDTGLIEDLGIPIRDMPLYKFPDKIKNDKDIAMHNKNIPYLN
EFMKGLARYSEKGLTRSTTSG"
6088..8053
/gene="C47F8.5"
/join(6088..6163,6222..6583,7339..7538,7593..7719,
7765..7854,7910..8053)
/gene="C47F8.5"
/codon_start=1
/protein_id="CA15838.1"
/db_xref="PID:e1344649"
/db_xref="PID:g3875100"
/db_xref="GI:3875100"
/db_xref="SPTREMBL:062114"
/translation="MFKILHINIOIKENLEFSKIKARGAOTTAERTMDYGSFITSFA
DIQTYKMLYLPETDIVRSPDILMNVASRTDSFARAVLRKTYMKNKYSILVDSGA
KALFLVGNVSEDYVRRTVMEAKIKGMVYIDEDTDDLPFKSLSLLAVAASAE
FKVLGKIDEDVFPDKLIPLDGKVIDPDAAAFYGLLKEGEPIYTKKDKAWYVPDY
AYNCTGPAYVAGFYLTRAAKVILFTFKQNFMTEDSLITGILANDIGIPKKNL
EHVYRHDYDIDNGEKEILAWHSEKNNIPYRNFVKVARRHYEQNLNMLEIYR"
complement(10441..11001)
/gene="C47F8.7"
/complement(join(10441..10661,10705..10797,10848..11001))
/gene="C47F8.7"
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CA15834.1"
/db_xref="PID:e1344645"
/db_xref="PID:g3875096"
/db_xref="GI:3875096"
/db_xref="SPTREMBL:062110"
/translation="MLPAPATFLILISIVAMHVOCPCPGFVNKTADYLCCKASKNV
ILVANSKDAIVETFEFTIKSYSESETPSDDDNNILPEKVAATSVATSVLPHTSKNI
TRIVKQALRGVGRFELIAEIPKHSRRPQVLYVICTIRHFSFIPLNDML"
complement(12091..13485)
/gene="C47F8.8"
/complement(join(12091..12144,12193..12267,12332..12533,
12584..12810,12859..13034,13207..13272,13317..13485))
/gene="C47F8.8"
/note="similar to zinc finger, C4 type (two domains)"
/codon_start=1
/protein_id="CA15840.1"
/db_xref="PID:e1344651"
/db_xref="PID:g3875102"
/db_xref="GI:3875102"
/db_xref="SPTREMBL:062116"
/translation="MGLISKRNKPGQVCHNTETTRHFGIISCTACAAFFRSISGMQY

```

```

ICTAENPCTISFDRKFRGACRYASCIKAKKLEHSHSPATSEGGSSSKTDESD
PEYSIKLMHDDLLQYIVKVKMAEKRRGCKENTQITITISDMLITTRQSPSK
EELCRPGVDLEKEDEVALRYRFSNIVDSVLISVLSAGKSDITDIDRLSR
IOEVKSTIGSSILAHNLNIEFAAFKALCIKWLKEFSAQCKIIAHFHYBGVGLAR
NYENCTELADMEIAIRIGDTITQIVTYQNTYHHEVRLYQIGMPL"
15076..17404
/gene="C47F8.2"
/join(15076..15091,15138..15365,16255..16831,17007..17208,
17252..17404)
/gene="C47F8.2"
/note="similar to zinc finger, C4 type (two domains)"
/codon_start=1
/protein_id="CA15836.1"
/db_xref="PID:e1344647"
/db_xref="PID:g3875098"
/db_xref="GI:3875098"
/db_xref="SPTREMBL:062112"
/translation="MSRSEFNSKICIGSICFAKPHRGISCNVACSEFRRTVSLNT
RYLKNNSNCTISPGQRTFCACRRESCVDAGAKRDPVOQRSTNATPKITLNTSG
SSNREVVAGLSISRTISTITFTGFTDSNHLVQIDHLNIDNNCHKMKKYSDL
LEISHMNLAREYVQVBEPEKELVSTLSNLAVLPQNIIOIKISINDIMTCAVNG
SAMESCVCPGTGKLSQEDVRLQKSFQFATWIDSTQYSIATDIDSESGESENT
SOFISHIKSTIGACLANIKLVDEFAAFKSCVMSIGQATGESKVVAAOHEEYVAS
AKMYVKTETSMDSSEIATRIEELSFGRSSVQMTYDITIKLYKIGIPKSLFPGN"
19133..20530
/gene="C47F8.1"
/join(19133..19209,19717..20340,20386..20530)
/gene="C47F8.1"
/note="predicted using GeneFinder: cDNA EST YK22967.5
comes from this gene; cDNA EST YK22967.3 comes from this
gene"
/codon_start=1
/protein_id="CA15835.1"
/db_xref="PID:e1344646"
/db_xref="PID:g3875097"
/db_xref="GI:3875097"
/db_xref="SPTREMBL:062111"
/translation="MSSFOLOQLPSLPKEVFKSLGVCDLNLICRSSNKSRLVYKSL
KDITFTLIVAEKHYKIMLKEDPSKHAARYRAENSNGTSLEKNIKRMHREVEYV
LDCTNVEGDVAVKYEYLLKKEPYLKLFECPMISDSLOMLNCEADVILISYDR
KVTFEDNVPILFNECITLHGKFRSEKGMNLFRTWOSKREKRTLOCVSLTRIGFPM
KQANEKYLKIDVSEYKRIKIPYKDKIKSTTYNLYKNDGELATIKFKQKEICFH
TENDR"
complement(613..679),complement(487..564),
complement(70..431),complement(281553.1.36172..36371),
complement(281553.1.36000..36126),
complement(281553.1.35864..35953),
complement(281553.1.35704..35817),
complement(281553.1.35704..35817)
/gene="C47F8.3"
/join(complement(613..679),complement(487..564),
complement(70..431),complement(281553.1.36172..36371),
complement(281553.1.36000..36126),
complement(281553.1.35864..35953),
complement(281553.1.35704..35817))
/gene="C47F8.3"
/codon_start=1
/protein_id="CA15841.1"
/db_xref="PID:e1344652"
/db_xref="PID:g3875103"

```

Query Match 75.7%: Score 17.4; DB 36; Length 21816;

Best Local Similarity 65.2%: Pred. No. 2.2e+02;

Matches 15: Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAARGCCRAARTCDATRTGRAA 23

Db 1174 GAATGCCAAATCCCTATGGAA 1152

RESULT 7

LOCUS HS992D9 155147 bp DNA PRI 23-NOV-1998

DEFINITION Human DNA sequence from BAC 992D9 on chromosome 22q12.1 contains


```

variation      /note="LM3 repeat: matches 909. .150 of consensus"
                9630. .9632
                /note="clone 90L6; GTC in this entry; substitution"
                /replace="ggc"
variation      /note="clone 90L6; CAG in this entry; substitution"
                /replace="cgg"
                10293. .10361
                /note="MIR repeat: matches 77. .144 of consensus"
                10337. .10339
variation      /note="clone 90L6; TCA in this entry; substitution"
                /replace="tta"
                11450. .11545
                /note="MIR repeat: matches 49. .145 of consensus"
                11956. .12000
                /note="MIR2 repeat: matches 146. .102 of consensus"
                12714. .12790
                /note="MIR2 repeat: matches 47. .124 of consensus"
                12903. .12936
                /note="U2 repeat: matches 34. .1 of consensus"
                13166. .13546
                /note="LM35 repeat: matches 535. .921 of consensus"
                13242. .13244
                /note="clone 90L6; ACA in this entry; substitution"
                /replace="ata"
                13656. .13658
variation      /note="clone 90L6; TGT in this entry; substitution"
                /replace="ttt"
                13916. .13918
variation      /note="clone 90L6; CAA in this entry; substitution"
                /replace="cca"
                14086. .14088
variation      /note="clone 90L6; AGA in this entry; substitution"
                /replace="aaa"
                14151. .14153
variation      /note="clone 90L6; GTG in this entry; substitution"
                /replace="ggg"
                14284. .14286
variation      /note="clone 90L6; GTG in this entry; substitution"
                /replace="ggc"
                14609. .14896
                /note="AluY repeat: matches 295. .7 of consensus"
                15373. .15666
                /note="AluX repeat: matches 293. .1 of consensus"
                15615. .15617
variation      /note="clone 90L6; CCT in this entry; substitution"
                /replace="ctt"
                15798. .15800
variation      /note="clone 90L6; AGG in this entry; substitution"
                /replace="aag"
                15814. .16224
                /note="MIR1B repeat: matches 390. .1 of consensus"

repeat_region

Query Match      74.88; Score 17.2; DB 9; Length 155147;
Best Local Similarity 70.08; Pred. No. 3e+02;
Matches 14; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY      4 RTGCCAARTCDATRTGAA 23
        :||||:||||:||||:||||:
Db 100531 ATGCCAAGTCATGTGCA 100550

RESULT      8
AF067844      AF067844      218336 bp      DNA      PRI      09-FEB-1999
LOCUS      *Homo sapiens chromosome 10 clone PTEN, complete sequence.
ACCESSION      AF067844
VERSION      94240386
KEYWORDS      AF067844.1 GI:4240386
SOURCE      human.
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

REFERENCE
AUTHORS      1 (bases 1 to 218336)
                Jensen,K., de la Bastide,M., Parsons,R., Parnell,L.D., Dedhia,N.,
                Gottesman,T., Gnoj,L., Kaplan,N., Lodhi,M., Johnson,A.F.,
                Shohdy,N., Hasegawa,A., Haberman,K., Huang,E.N., Schuttz,K.,
                Calma,C., Granat,S., Wigler,M. and McCombie,W.R.
                Genomic sequence of PTEN/MMAC1
TITLE      2 (bases 1 to 218336)
JOURNAL      Unpublished
REFERENCE      Jensen,K., de la Bastide,M., Parsons,R., Parnell,L.D., Dedhia,N.,
AUTHORS      Gottesman,T., Gnoj,L., Kaplan,N., Lodhi,M., Johnson,A.F.,
                Shohdy,N., Hasegawa,A., Haberman,K., Huang,E.N., Schuttz,K.,
                Calma,C., Granat,S., Wigler,M. and McCombie,W.R.
                Direct Submission
TITLE      Submitted (18-MAY-1998) Lita Annenberg Hazen Genome Sequencing
JOURNAL      Center, Cold Spring Harbor Laboratory, 1 Bungtown Rd., Cold Spring
                Harbor, NY 11724, USA
FEATURES
SOURCE      1. .218336
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="10"
                /clone="PTEN"
                1. .106991
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="BAC 265N13"
                /chromosome="10"
                22308. .22338
                /gene="PTEN"
                /note="5'-UTR defined by comparison to PTEN cDNA U93051"
                join(22308. .23417,51995. .52079,83482. .83526,89015. .89058,
                90987. .91225,110086. .110227,115821. .115987,118862. .119086,
                123258. .124345)
                /gene="PTEN"
                /note="mRNA coordinates delineated by comparison to PTEN
                cDNA U93051"
                22308. .124345
                /gene="PTEN"
                /note="the coding region of PTEN, as defined by the cDNA,
                identifies 9 exons within this region; identical to MMAC1
                (U92346) and PTEN (U93051)"
                /evidence="experimental"
                22308. .23417
                /gene="PTEN"
                /function="5'-UTR and initial segment of the CDS"
                /number=1
                join(22339. .23417,51995. .52079,83482. .83526,89015. .89058,
                90987. .91225,110086. .110227,115821. .115987,118862. .119086,
                123258. .123443)
                /gene="PTEN"
                /note="coding regions delineated by comparison to PTEN
                cDNA"
                /codon_start=1
                /product="PTEN"
                /protein_id="AA013528.1"
                /db_xref="PID:94240387"
                /db_xref="GI:4240387"
                /translation="MTAIIKEIYSRNRRROEGFDLDTIYIPNITAMGFPARLEG
                VYRNNDIDVVRFLDSKRHNKIKYINLCAEHRYTDARKNCVAVQYPPEDHNPQDELK
                PFCEDLDQWLSEDDNHVAATHCKAGKRTGVMICAYLLHGRKFLKQDALDFGEVVT
                RDKKGVIPDSRRVYVYSLIKNHLDRVALLHFMKEFIIPMSGGTCNPFQVVC
                OLKVKIYSNSGPTRRDREKMFPEFPOPLFVCGDIKVEFFHKKRMKDKKMFHWYV
                TFIPIGSEETSEKVENSLCDOELDSCSTERADNCKEYVLLITKNDLKKAKDKAN
                RYSPNKKVILYFTKYVEPSNPDASSSTVTPDVSNPDHRYISDTTSDPENEPF
                DEQHTQITKV"
                51995. .52079
                /gene="PTEN"
                /number=2
                58169. .218336
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="BAC 60C5"

```


NNNVNLKQHEDELLIVKKNVQIQONPKRGVSLILIIEMPNLNNRVEYVAAEQIRI*
 AMPFASLEKHTYIKLNNVYPKOTDELKKAELAHILDEKMGIRKEV*
 complement(7643. .9025)
 /gene="lpa-8r"
 complement(7643. .9010)
 /gene="lpa-8r"
 /codon_start=1
 /transl_table=11
 /protein_id="CA51564.1"
 /db_xref="PID:9413932"
 /db_xref="GI:413932"
 /db_xref="SWISS-PROT:P25152"
 /translation="MKKLLVNMNAVITATLILLPASVYPAHAHVQINSERELPRK
 AKRAYSTISOLSAIGPRTGRLTAERKSAALLIISNMKLLDVAQVQFNIPDRLEGL
 SSGARVTLQASGASAPTEBOGLTAPLYNGLGYQDPTADAGKIALISRGDLYTE
 KANNAEAGAKAVIIVNNKESLVPNTPNLSGNGVGLPVVGIKEDGALQOKEATLK
 LKPFNTSONIIGIKKPKIKHPDIYYVAHDSVPSGADNGSGTVMLEAVL
 LKVPSPDSNRTFYAPGAEEELGLGSSHYVHLSERKEKREVEVNLDMVGTSMKASE
 LYVNTLDGQSNRYWESSRTYAEKIRQDPSLTLTGCGSSDRHPFHEAGDLSANFTMGDE
 TEVEVPIHTPEDSIEHTIKERIKQADGLTAAVIAVEAKNKKPKITIKOMKANASDI
 FEDIK"
 complement(9020. .9025)
 /gene="lpa-8r"

Query Match 73.9% Score 17; DB 1; Length 97015;
 Best Local Similarity 65.2% Pred. No. 3/7e+02;
 Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

1 RAARTGCCBARATCDATRTGRGA 23
 Db 21697 GAARTGCCGAARATCGTTTGGAA 21675
 :||:||||:||||:||||:|

RESULT 10
 LOCUS BSUB0020
 DEFINITION Bacillus subtilis complete genome (section 20 of 21): from 3798401
 to 4010350.
 ACCESSION 299123 AL009126
 NID 92636240
 VERSION 299123.1 GI:2636240
 KEYWORDS
 ORGANISM Bacillus subtilis.
 SOURCE Bacillus subtilis.
 ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae; Bacillus.

1 (bases 1 to 212150)
 Kunst,F., Ogasaara,N., Moszer,I., Albertini,A.M., Alloni,G.,
 Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
 Boriss,R., Boursier,L., Brans,A., Braum,M., Brignell,S.C.,
 Bron,S., Brouillet,S., Brusch,C.V., Caldwell,B., Capuano,V.,
 Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
 Daniel,R.A., Denizot,F., Devline,K.M., Dusterhoft,A., Ehrlich,S.D.,
 Emerson,P.S., Enlian,K.D., Errington,J., Faderl,C., Ferrali,E.,
 Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
 Galleon,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
 Grandi,G., Giesepf,H., Guy,B.J., Haga,K., Halech,J., Harwood,C.R.,
 Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
 Itaya,M., Jones,L., Joris,B., Karamata,D., Kaahara,Y.,
 Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
 Konigstein,G., Krogh,S., Kunano,M., Kurita,K., Lapidus,A.,
 Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
 Masuda,S., Mauel,C., Medigue,C., Medina,N., Melindo,R.P.,
 Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
 Ogawa,K., Ogikawa,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
 Portellelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
 Purnelle,C., Rocha,E., Roche,B., Rose,M., Sadate,Y., Sato,T.,
 Scanlan,B., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
 Sekowska,A., Seror,S.J., Serro,P., Shin,B.S., Soldo,B.,
 Sorokin,A., Tacconi,E., Takai,T., Takahashi,H., Takematu,K.,
 Takeuchi,M., Yamakoshi,A., Tanaka,T., Tenister,P., Toyonoli,A.,
 Tosato,V., Uchiyama,S., Vandenberg,M., Vannier,F., Vassartotti,A.,

TITLE
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
NATURE 390 (6657), 249-256 (1997)

JOURNAL
98044033

MEDLINE
2 (bases 1 to 212150)

REFERENCE
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submision
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
danchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES
source
1. 212150
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
complement(7. .991)
/gene="ywkB"
complement(7. .23)
/gene="ywkB"
complement(32. .991)
/gene="ywkB"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15732.1"
/db_xref="PID:e1186205"
/db_xref="PID:g2636241"
/db_xref="GI:2636241"
/db_xref="SWISS-PROT:P45869"
/translation="MSLDILFLAPIFEVIVLGMFAGHSGDAKSGVSTLVYKY
ALPAFAGILTSRSEPSQPLMSILIGVGYITILLRCIFRYKDLNVSFV
LNSAQPTFAKGIPIVGLFCANEVAIPAVGIYVNLIDPLALITIVAGSSKNE
EGSDSEFKWTKGKSLHGLCEPLAAPLSMILVFNFTLPELGVKMDQLSTTSV
GSAVAAGVIGIKRILSMALGIALKVAVCPALMFLALALGLPADDTTKALLVAF
PGSAVAMATATFEKEETATAFVSALLSLISPIITALTA"
complement(1073. .2821)
/gene="ywkA"
complement(1073. .2821)
/gene="ywkA"
/function="unknown"
/note="similar to malate dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB15733.1"
/db_xref="PID:e1186206"
/db_xref="PID:g2636242"
/db_xref="GI:2636242"
/db_xref="SWISS-PROT:P45868"
/translation="WGYLTWTTISRKKEICLNIRKKEGLETTLAGKEVLSIPTL
KGVAFSLBEROELBGLPPTVSLDQQAQRAEQFOAPDRLRQNVYISDLNKN
EVLFTKLLKNLREMLPVYITPTGEAAQESHERRPGIYLSIDNDIGEKAVENL
HATAGDIDLIVATDSBSILGIDMGVGINIIGRLAYTAAGIDPSRVIPVVDVG
TNNERLNDPLIYGNKHERVOGEREAFIDAYVKAALKEFPALLHMDLGNKNRNI
MKKYNEHLTENDDIQGTATLAVIAMKRTGASINDORVITGASAGIGINDOI
RDTVYLAGLSESEANKRPETLDYRGLNEDIGITDFPKPIRNADPYKDKRDEKGO
IPDEYVRAKFTIILIGTSGVSGATFEEVYKEMASHVDPPVIMPSNPTHLAEAVPD
LFRWTDGVALLTGSPPDNVENYSGSYEIGSNMFAFPGGLGSLVNAEARIITPAMF
AATADAIMEVDLETGAGLLPESIDKLQEVSIQVAIAVAEAIKQGVANRPEDYKQA
VLDAWTPREYKVIK"
complement(3011. .3647)
/gene="tdk"
complement(3011. .3031)
/gene="tdk"
complement(3060. .3647)
/gene="tdk"
/EC_number="2.7.1.21"

/codon_start=1
/transl_table=11
/product="thymidine kinase"
/protein_id="CAB15734.1"
/db_xref="PID:e1186207"
/db_xref="PID:g2636243"
/db_xref="GI:2636243"
/db_xref="SWISS-PROT:Q03221"
/translation="WYIMKQSGWLELTCGSMFSGKSEELLIRNRKATYANQOEVRFP
VIDRKYSAAVAVSNGTSMSTSYAISADIDWDHISESTDVAVADVEQFEOELVEVLS
SLADKGVYVLAGLMDPDRGEPFGVFNIMAIASVTKYLAQVAVCSGSPASRTQLRID
GKPSYDDPVTLVGAASYEARCCHHHEVGRKSK"
complement(3698. .3936)
/gene="rpmB"
complement(3698. .3720)
/gene="rpmB"
complement(3736. .3936)
/gene="rpmE"
/codon_start=1
/transl_table=11
/product="ribosomal protein L31"
/protein_id="CAB15735.1"
/db_xref="PID:e1186208"
/db_xref="PID:g2636244"
/db_xref="GI:2636244"
/db_xref="SWISS-PROT:Q03223"
/translation="MKAGIHPNFKATVKAQGNFETGSVKEEVRVEICSECHPFT
GRQKFSADGKVDYDRFNKYGK"
complement(4055. .5338)
/gene="rho"
complement(4055. .5338)
/gene="rho"
/codon_start=1
/transl_table=11
/product="transcriptional terminator Rho"
/protein_id="CAB15736.1"
/db_xref="PID:e1186209"
/db_xref="PID:g2636245"
/db_xref="GI:2636245"
/db_xref="SWISS-PROT:Q03222"
/translation="MKDVYSISLENNMLKELYELARHYKISYSGKLRKELIFALNA
NABOEDLMEGVLEILOSBEQFRLPINSPESEDIYISAOIRFEDLRNGDVSGK
VRPKENRERYGILHFAVNVGDDESAKREYHPALPLIPDRQMTLETDPNLSPTI
MDMAPIYVGQRCGLIVAPPKAKGTMLKEIANSITNAPAEELIYLLIDEREPEYTI
ERSVAGDVSTFDEVPENHIVALELVERAMRVERKVTYIIMOSITRLARVNIY
IPPSGRILSGIDPAAPRRPRFPGARNTIEGGSYLTLATVDTGSRMDVYEEF
KGTGNMELHLDRLSLAERRIFPAIDIRRSGRKEELVLVPEKHLDRIMSIKRTWSDSPDF
AEKFMKMKKTKTQGEFFDILNMQKQANASAR"
complement(5714. .5735)
/gene="ywjA"
complement(5714. .6710)
/gene="ywjA"
complement(5745. .6710)
/gene="ywjI"
/function="unknown"
/note="similar to glycerol-inducible protein"
/codon_start=1
/transl_table=11
/protein_id="CAB15737.1"
/db_xref="PID:e1186210"
/db_xref="PID:g2636246"
/db_xref="GI:2636246"
/db_xref="SWISS-PROT:Q03224"
/translation="MERSLSMELVRVVEAALASARMGKGRKDEADEAATSAARDVF
DYVMKGTIVYGEEMDEAPMLYIGELKNGVGRVAVDPLEGTNIIASGKNALT
VIAVADGTLNAPDMYMNKAVGPEAVGCIIEAPYDNLKAVAKAKKNDDEVYVAT
TINERAKTISERREAGARIKINDQVAGAITAPADRHGVLDLSESGAPBEVLSA
VALKALGELIIGKLPOSEBITRCHRWGLDLSLVLMEDLVKQDDHIFAATGTGDE
LLKGVQFGSVGTIESLVIIRAKSGTVFVDGRSLKRRKPLVIRP"
complement(6741. .8030)
/gene="murZ"
complement(6741. .8030)
/gene="murZ"

/EC number="2.5.1.7"
/function="peptidoglycan biosynthesis"
/note="alternate gene name: murZ, lssF, rev-4"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine
1-carboxylinyltransferase"
/protein_id="CA85738.1"
/db_xref="PID:e1186211"
/db_xref="PID:g2636247"
/db_xref="GI:2636247"
/translation="MERKTAGDPSLNGTVHISGAKKSAVALPATILANSEYITIEL
PISIDITLIDKKEIGNHFENGEMVNDPTSMISPLNGYKKRIRASGYMGAL
GRKQAVIGPSCGCHLPRIIDHIFKFEALGAVTTEOGATYIRARLGRAGTIDY
VSVGATIMLAVALAGKTTIENNAKEPITIDVATLTSMNAIKGAGTNGVIRIDV
KELHGGKHTIIPRIENAGTFMIAAGKREVIIDNVIPTHLESTALTRMGYHIEHS

Query Match 73.9%; Score 17; DB 1; Length 212150;
Best Local Similarity 65.2%; Pred. No. 3.8e+02;
Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAARTGCCRAARTCDATRTGAA 23
:|||||:|||||:|||||:
Db 136262 GAATGCCGAATCGTTTGAA 136284

RESULT 11
LOCUS YSCMET10A 3650 bp DNA PLN 08-FEB-1995
DEFINITION Saccharomyces carlsbergensis assimilatory sulfite reductase (MET10)
ACCESSION L26504
NID 9498310
VERSION L26504.1 GI:498310
KEYWORDS MET10 gene.
SOURCE Saccharomyces carlsbergensis DNA.
ORGANISM Saccharomyces pastorianus
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 3650)
Hansen, J., Cherest, H. and Kleveland-Brandt, M.C.
Two divergent MET10 genes, one from Saccharomyces cerevisiae and
one from Saccharomyces carlsbergensis, encode the alpha subunit of
sulfite reductase and specify potential binding sites for FAD and
NADPH.
Mol. Cell. Biol. 13:176 (1993), 6050-6058 (1994)
COMMENT
MIDLINE J. Bacteriol. 176 (19), 6050-6058 (1994)
REFERENCE 2 (bases 1 to 3650)
Hansen, J.
Direct Submission
Submitted (03-DEC-1993) J. Hansen, Carlsberg Research Laboratory,
Gamle Carlsbergvej 10, Copenhagen Valby, Denmark, 2500
On Jun 10, 1994 this sequence version replaced gi:432489.
FEATURES
source
1. 3650
location/Qualifiers
/organism="Saccharomyces pastorianus"
/db_xref="taxon:27292"
215..3322
/gene="MET10"
215..3322
/EC_number="1.8.1.2"
/note="flavin-binding subunit"
/codon_start=1
/product="assimilatory sulfite reductase"
/protein_id="AA61981.1"
/db_xref="PID:g432480"
/db_xref="GI:432490"
/translation="MSVEFATNPFQGTAKATLPSYTPVATISSVLEFNNVDSIFAK
SFSQADLHDEKLMKMSLQNSNGSKATFQELDRSAGLPIGFSGLNNITAVP
GSLAPYFINSIQIPVHAKFLINAGALINTSGSVINDYVATDAASKLNITAVP
STNEOVALLALAVATFNSGAINFDGANYSKTVPVAVETAAESSIIIGKLSKVA
PNAFEDVLKFNLSGLRLHNPQIFPAQDAETVFIYGSLESELSISGNNAKIG

BASE COUNT 1101 a 743 c 696 g 1110 t
ORIGIN

Query Match 73.9%; Score 17; DB 7; Length 3650;
Best Local Similarity 65.2%; Pred. No. 3.2e+02;
Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAARTGCCRAARTCDATRTGAA 23
:|||||:|||||:|||||:
Db 2247 GAATGCCAATTCATGTGAA 2225

RESULT 12
LOCUS HS19408
DEFINITION Human DNA sequence from clone 19408 on chromosome 6q24.1-25.3
ACCESSION AL031769
NID 94153934
VERSION AL031769.1 GI:4153934
KEYWORDS HNG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 126839)
Smith, S.
Direct Submission
Submitted (18-MAR-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jan 13, 1999 this sequence version replaced gi:4007130.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence is the entire insert of clone 19408. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unseq'
feature key
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre Chromosome 6
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
19408 is from the library RpgII constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR:PCRAC2.
FEATURES
source
1. 126839
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="19408"

```
repeat_region /note="q24.1-25.3"
/clone_lib="RPC11"
1. .100
repeat_region /note="Aluub repeat: matches 194. .293 of consensus"
224. .431
/note="L2 repeat: matches 2279. .2496 of consensus"
2543. .3170
/note="LIMC repeat: matches 410. .1067 of consensus"
3261. .3461
/note="L2 repeat: matches 2352. .2539 of consensus"
5020. .5314
/note="Aluy repeat: matches 1. .295 of consensus"
5894. .6199
/note="LIM47 repeat: matches 5982. .6289 of consensus"
6200. .6733
/note="LIP415 repeat: matches 5595. .6144 of consensus"
6952. .7260
/note="Alusx repeat: matches 6. .303 of consensus"
7439. .8958
/note="LIM4 repeat: matches 2116. .3733 of consensus"
9186. .10885
/note="LIMB5 repeat: matches 4442. .6173 of consensus"
11407. .11518
/note="MLTID repeat: matches 1. .107 of consensus"
11520. .11640
/note="MST-INTERNAL repeat: matches 1520. .1650 of
consensus"
11681. .11934
/note="MLT1-INTERNAL repeat: matches 964. .1212 of
consensus"
12067. .12284
/note="MLT1-INTERNAL repeat: matches 657. .889 of
consensus"
12513. .13673
/note="Tiggrerb repeat: matches 109. .1231 of consensus"
13674. .13986
/note="Alusg repeat: matches 1. .313 of consensus"
13987. .14017
/note="Tiggrerb repeat: matches 1. .109 of consensus"
14019. .14109
/note="MLT1-INTERNAL repeat: matches 405. .503 of
consensus"
14116. .14211
/note="MST-INTERNAL repeat: matches 201. .293 of consensus"
14247. .14469
/note="MLTID-internal repeat: matches 4. .225 of consensus"
14472. .14810
/note="MLTID repeat: matches 164. .505 of consensus"
15542. .16687
/note="L2 repeat: matches 2602. .2745 of consensus"
16724. .16792
/note="MIR repeat: matches 160. .233 of consensus"
16875. .17286
/note="L2 repeat: matches 1774. .2234 of consensus"
17606. .17975
/note="THEIC repeat: matches 1. .371 of consensus"
18633. .18864
/note="MIR repeat: matches 9. .261 of consensus"
19203. .19393
/note="MLT1H repeat: matches 339. .525 of consensus"
19992. .20371
/note="L2 repeat: matches 2318. .2749 of consensus"
20461. .20585
/note="L2 repeat: matches 2541. .2667 of consensus"
21143. .21217
/note="MIR repeat: matches 25. .108 of consensus"
21555. .21850
/note="Aluud repeat: matches 5. .289 of consensus"
22343. .22761
/note="LIM4 repeat: matches 1492. .1907 of consensus"
22762. .23474
/note="LIM4 repeat: matches 2732. .3507 of consensus"
23475. .23618
repeat_region /note="Aluy repeat: matches 1. .144 of consensus"
23681. .23985
/note="Alusx repeat: matches 1. .297 of consensus"
24112. .24313
/note="MIR repeat: matches 41. .251 of consensus"
24716. .25220
/note="WER77 repeat: matches 57. .626 of consensus"
25365. .25660
/note="Aluub repeat: matches 1. .293 of consensus"
26465. .26504
/note="L20 copies 2 mer ct 80% conserved"
26924. .26969
/note="L23 copies 2 mer ag 87% conserved"
27427. .27603
/note="Aluy repeat: matches 128. .304 of consensus"
31176. .31483
/note="Alusg repeat: matches 1. .307 of consensus"
31485. .31638
/note="L77 copies 2 mer aa 68% conserved"
31485. .31629
/note="L5 copies 29 mer 71% conserved"
31905. .32424
/note="match: GSS A0312911 clone R-1A24"
32639. .32958
/note="WER18 repeat: matches 1. .337 of consensus"
33466. .33713
/note="L1 repeat: matches 4642. .4906 of consensus"
33909. .33987
/note="L2 repeat: matches 2639. .2748 of consensus"
34449. .34527
/note="LIP412 repeat: matches 5562. .5640 of consensus"
34531. .34728
/note="MLT2 repeat: matches 1. .214 of consensus"
34845. .35074
/note="MLT2A repeat: matches 211. .453 of consensus"
35075. .35593
/note="LIP412 repeat: matches 5639. .6165 of consensus"
36282. .36570
/note="Alusg repeat: matches 1. .289 of consensus"
36599. .37953
/note="LIP416 repeat: matches 4781. .6150 of consensus"
/note="complement(<37973. >38339)
/note="match: GSS A0180570"
39802. .39867
/note="L33 copies 2 mer ta 67% conserved"
40356. .40425
/note="L35 copies 2 mer tt 70% conserved"
40611. .40686
/note="L38 copies 2 mer tt 66% conserved"
41665. .42458
/note="LIP47 repeat: matches 5293. .6143 of consensus"
42488. .42519
/note="L16 copies 2 mer at 94% conserved"
42521. .43181
/note="LIMB1 repeat: matches 5491. .6170 of consensus"
/note="complement(42981. .43420)
/note="match: 278716 chromosome 6 HindIII fragment"
43796. .43949
/note="FRAM repeat: matches 1. .154 of consensus"
44538. .44908
/note="MLT2CA repeat: matches 1. .452 of consensus"
44909. .44995
/note="L3 copies 29 mer 79% conserved"
44973. .45002
/note="L15 copies 2 mer ac 87% conserved"
46196. .46355
/note="MIR repeat: matches 47. .227 of consensus"
46558. .46916
/note="L2 repeat: matches 975. .1346 of consensus"
49411. .49699
/note="Alusg repeat: matches 1. .289 of consensus"
49860. .50123
/note="Alusg1 repeat: matches 1. .285 of consensus"
```


RESULT 15
 CELB0207 43584 bp DNA INV 27-MAY-1999
 LOCUS
 DEFINITION *Caenorhabditis elegans* cosmid B0207.
 ACCESSION U97196
 NID 91943799
 VERSION 097196.1 GI:1943799
 KEYWORDS
 SOURCE
 ORGANISM
Caenorhabditis elegans.
Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; *Caenorhabditis*.
 1 (bases 1 to 43584)
 The *C. elegans* Genome Sequencing Consortium, Washington University
 Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
 Hinxton, U.K., C.
 TITLE
 Genome sequencing of the nematode *C. elegans*: a platform for
 investigating biology. The *C. elegans* Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 JOURNAL
 MEDLINE
 99069613
 ERATUM: [[published erratum appears in Science 1999 Jan
 1;283(5398):35]]
 2 (bases 1 to 43584)
 Henhaus, J. and Woldmann, P.
 TITLE
 The sequence of *C. elegans* cosmid B0207
 JOURNAL
 REFERENCE
 3 (bases 1 to 43584)
 Waterston, R.
 TITLE
 Direct Submission
 JOURNAL
 REFERENCE
 4 (bases 1 to 43584)
 Waterston, R.
 TITLE
 Direct Submission
 JOURNAL
 REFERENCE
 5 (bases 1 to 43584)
 Submitted (27-MAY-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 e-mail: r.w@genome.wustl.edu and j.s@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one m13 subclone.

NEIGHBORING COSMID INFORMATION:

The 5' clone is Y76G2B, 3950 bp overlap; 3' clone is Y76G2A, 200 bp
 overlap. Actual start of this clone is at base position 1 of
 CELB0207; actual end is at 43584 of CELB0207

NOTES:

Coding sequences below are predicted from computer analysis, using
 the program GeneIndexer P. Green and L. Hillier, ms in preparation).

FEATURES

1..43584
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"

gene
 CDS

gene
 CDS

gene

/db_xref="taxon:6239"
 /clone="B0207"
 /chromosome="I"
 44..3900
 /gene="B0207.5"
 join(44..2693,2739..2873,3125..3204,3775..3900)
 /note="coded for by C. elegans cDNA yk21b6.5; coded for by
 C. elegans cDNA yk5f10.3; coded for by C. elegans cDNA
 yk47d12.3; coded for by C. elegans cDNA yk130c11.3; coded
 for by C. elegans cDNA yk47d12.5; coded for by C. elegans
 cDNA yk130c11.5"
 /codon_start=1
 /protein_id="AAB52460.1"
 /db_xref="PID:91943806"
 /db_xref="GI:1943806"
 /translation="MSTTTTTPAEEYVYDEEYEVITLVKESRRPLKYSR
 PDDEDDDEVYREKPKSYTFEGFVYRGPRKIRKRPSTDSGEPPDIKRYRKY
 EGEEDRGVLRKSHKRLIVRMKDEEPEEHEVRVYKAKWKGEKNEPND
 ENVDGKTIIVRNRRQYFERNDPRDDEEELVNRVAVRGAAPKIFNAHPRGR
 SAKPSITERNSEEFYKIGSHEENHRKKNQKREDYRRRRKPEYEESEELGR
 RKPRGESEEFERKSGRPDREONGETTVIKYSGSISDPTPLPDQGESVAITK
 KIKFYGKERDTPDDEDENEIKILKRLVIGARDKPTGNDNGEIVGLITTKKSEF
 ICPNGKQPIYGNDEDEPFRKRTSDHPRKODEVOLFKKNTFRIVNGKVLRTLE
 DIPQDEDESVKLTKKKRYRIGQNGEITDEDESRSRKISDEPEDEDOIILP
 KRKFYKIVDGRKVLITGEGVPQGGENGEOIKTKKRYRIGNGEITKSSDDOAL
 PSHDRGKLVIGKNNRLYRKGEIPEDEENGESYVVMKRYQLEKPOSPDEEEDER
 FKLLTKKKKGVIANPDNSGEIHKMKKAYLVIGESMDQDESEEMMKR
 RYKVGADGKREILREBEIPMGOLGTSKSSYSPPLDGERIEKTRIKYIKGRDGOEL
 OKMGDDESGDSQGTGTHYVMKRRKTRRYANSEKELIKENYPGDLSGDEBDGKA
 IWSKTKRSTYIDGKNNKLVKEKVRGEPEDDQOKLMGKTKRTRTKWYDGRKLM
 KTESDGDNNMPPKHNSTPNDGVPPPGHFSYVRSDDNNKRDGDSERGSDETR
 TRNGKTNKRDGRDGRNMLGSIIPKRRSSNDGNDSDNNDRREKTRSSRARSS
 RKMEKYHKDRNVRSSGSEELVPE"
 complement(4089..7564)
 /gene="B0207.6"
 complement(join(4089..4429,4757..4955,5113..5220,
 5338..5509,7482..7564))
 /gene="B0207.6"
 /note="weak similarity to ATPases"
 /codon_start=1
 /evidence-not_experimental
 /protein_id="AAB52461.1"
 /db_xref="PID:91943807"
 /db_xref="GI:1943807"
 /translation="MOMLANEKKOLLRVYATSEKGENLGSTHTMYGVLTGAPGA
 GKSTECAGIDITFEQKRPETLINDPRANDMAYAPDNTITEMVNDYMRRLGPN
 GALKICETLGCANCMLOKTEAHKKLILIDCGQQLIYSEBELKVIYTFLEKGV
 RLCAHLADSLYCSDPKSFISVALSTLATVTMTMPQVNCISKADLESDGTLDLEF
 SHLPVNRLLDLNLEPGLERYKRLNEAICGVISDFLVSEFPLAVENKESMRYIOM
 VDAANGFSLTEOGGIRELILNSKA"
 5897..7241
 /gene="B0207.4"
 join(5897..5961,6029..6103,6153..6261,6338..6705,
 6769..7042,7143..7241)
 /gene="B0207.4"
 /note="Contains similarity to Pfam domain: PF00069
 (pklnase). Score=295.1, E-value=2.8e-85, N=1; coded for by
 C. elegans cDNA CEES128f; coded for by C. elegans cDNA
 CEES194f; coded for by C. elegans cDNA CEES136f; coded for
 by C. elegans cDNA yk11a10.5"
 /codon_start=1
 /protein_id="AAB52459.1"
 /db_xref="PID:91943805"
 /db_xref="GI:1943805"
 /translation="MSTKNNALIELINSITKFIYISDMENKPPYINLPKETYNTPO
 KGRFTINDEFIIGPLGKRGSVYLAATKRGHVAATLVKSQLISGVEHOLERE
 IEIOSLHNPMILKITYFWDAKRIYVLEAPGGEMKQTVSKRSEEPNAAIYME
 IADLSYCHRKNIHRIIDPENLLISGGEILKIDGFSVAAIKARFTGTMDIYLP
 PEMVNGADHSADVDAIAGVLYCEYLVCKPPEHEDGQSTVAAIKARFTYDVKVG
 ARDLIGRLVYDPAKRLTLEQYKEHYVIGMMKAKIRAKQOKTEKASLANH"
 11129..114089
 /gene="gpa-14"

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:37 ; Search time 255.05 Seconds

(without alignments)
22.562 Million cell updates/sec

Title: US-09-325-095-29

Perfect score: 23

Sequence: 1 RAARTGCCRAARTCDARTGRMA 23

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73.9	9789	1	T41852	CDNA encoding plas
2	73.9	6831	1	T80200	Phosphatidyl inositol
3	72.2	38	1	OS9013	Primer for p110 CD
4	72.2	38	1	OS7019	Ptdins 3-kinase pr
5	72.2	1610	1	X04612	Genomic sequence e
6	69.6	1131	1	O48032	Flnd DNA. Vaccines
7	69.6	641	1	V75786	Staphylococcus aur
8	68.7	40875	1	T80043	Insert from cosmid
9	67.8	2540	1	O29627	Hepatitis C virus
10	67.8	2540	1	O43888	NABV hepatitis vir
11	67.8	2401	1	O48539	Bacillus subtilis
12	67.8	381	1	O57522	Human P17R-f CDNA.
13	67.8	393	1	O57523	Human P17R-f CDNA.
14	67.8	24	1	O59023	Degenerate primer
15	67.8	2540	1	O63752	NABV genomic fra
16	67.8	9502	1	O74770	Hepatitis C virus
17	67.8	1835	1	T29773	Maize adenylosucci
18	67.8	1835	1	T32829	Maize adenylosucci
19	67.8	3141	1	T71746	Fltrobacter succin
20	67.8	3141	1	V16533	CDNA encoding mam
21	67.8	5220	1	V31340	Human phosphatidyl
22	67.8	3549	1	X13796	Enterococcus faeca
23	67.8	8921	1	X13005	Enterococcus faeca
24	67.8	537	1	X13917	Enterococcus faeca
25	67.8	5220	1	X15932	CDNA sequence enco
26	67.8	1835	1	X25765	Maize adenylosucci
27	67.8	1946	1	N70526	Sequence of CDNA 1
28	67.0	11000	1	T58840	Continuation (4 of
29	67.0	997	1	V15021	652 strain leukoto
30	67.0	992	1	V15030	652 strain leukoto
31	67.0	5285	1	T80199	Phosphatidyl inositol
32	67.0	3140	1	V52291	Streptococcus pneu
33	67.0	5061	1	V42920	CDNA encoding a hu
34	67.0	32768	1	X13060	Enterococcus faeca
35	67.0	995	1	X13572	Enterococcus faeca
36	67.0	529	1	X30828	Streptococcus pneu
37	67.0	992	1	X28835	A.actinomycetemcom
38	66.1	2031	1	V59655	Human secreted pro
39	65.2	2111	1	O63889	Sequence coding de
40	65.2	2774	1	O66579	Sorbiol dehydroge
41	65.2	2664	1	O78644	E.coli pyruvate de
42	65.2	8041	1	T10105	Adherence conferr
43	65.2	3707	1	O99277	Human neurotrophic

ALIGNMENTS

C	44	15	65.2	3194	1	T00689
	45	15	65.2	35515	1	X20252
Human trkB recepto						
Borrelia burgdorferi						
RESULTS						
RESULT	1					
ID	T41852					
AC	T41852					
DE	20-FEB-1997 (first entry)					
KW	CDNA encoding Plasmodium falciparum erythrocyte membrane protein.					
KW	Plasmodium falciparum; erythrocyte membrane protein; malaria;					
OS	detection; identification; treatment; prevention; parasite; ss.					
OS	Plasmodium falciparum MC type.					
PH	Location/Qualifiers					
FT	Key					
FT	cds					
FT	326..9497					
FT	/*tag= a					
FT	/product= Erythrocyte membrane protein					
FT	518..520					
FT	/*tag= b					
FT	/transl_except= GAA encodes Tyrosine					
FT	656..658					
FT	/*tag= c					
FT	/transl_except= ATT encodes Leucine					
FT	2909..2911					
FT	/*tag= d					
FT	/transl_except= AAC encodes Aspartic acid					
FT	3461..3463					
FT	/*tag= e					
FT	/transl_except= GAA encodes Glutamine					
FT	5546..5548					
FT	/*tag= f					
FT	/transl_except= CCT encodes Arginine					
FT	6254..6256					
FT	/*tag= g					
FT	/transl_except= AAT encodes Lysine					
FT	6257..6259					
FT	/*tag= h					
FT	/transl_except= ATA encodes Tyrosine					
FT	6263..6265					
FT	/*tag= i					
FT	/transl_except= AAC encodes Lysine					
FT	6269..6271					
FT	/*tag= j					
FT	/transl_except= TTC encodes Isoleucine					
FT	6272..6274					
FT	/*tag= k					
FT	/transl_except= ATA encodes Histidine					
FT	6275..6277					
FT	/*tag= l					
FT	/transl_except= ATR encodes Asparagine					
FT	6278..6280					
FT	/*tag= m					
FT	/transl_except= GGA encodes Tryptophan					
FT	7754..8478					
FT	/*tag= n					
PN	intron					
PN	MO9633736-AL.					
PD	31-OCT-1996.					
PF	26-APR-1996; U05798.					
PR	27-APR-1995; US-430908.					
PA	(AFVY-) AFFYMAX TECHNOLOGIES NV.					
PI	Baruch DI, Howard RJ, Pasloske BL;					
DR	WPI: 96-497376/49.					
DR	P-PCDB: W00384.					
PT	New Plasmodium falciparum erythrocyte membrane proteins - used to					
PT	develop products for the diagnosis, treatment or prevention of					
PT	malaria parasite infections					
PS	Disclosure: Figure 12: 149pp; English.					
CC	A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte					
CC	membrane protein 1 (PfEMP1) or active fragments or analogues of that					
CC	protein can be used in the treatment or prevention of symptoms of a					

CC malaria parasite infection. The polypeptides can inhibit, block or
 CC reverse the sequestration of erythrocytes in patients suffering from
 CC malaria. Nucleic acids derived from the PfEMP1 gene can be used as
 CC probes and primers to identify a Plasmodium falciparum parasite, the
 CC primers used to generate characteristic amplification patterns from
 CC different P. falciparum strains. Antibodies specifically
 CC immunoreactive with the PfEMP1 polypeptide or its fragments may be
 CC used in diagnosis of malaria infection. This sequence encodes the
 CC PfEMP1 protein of the MC type of Plasmodium falciparum. An
 CC alternative, truncated version of the coding sequence (a cDNA clone)
 CC is given in T41853.
 SQ Sequence 9789 BP; 4061 A; 1393 C; 1837 G; 2498 T;

Query Match 73.9%; Score 17; DB 1; Length 9789;
 Best Local Similarity 65.2%; Pred. No. 25;
 Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

OY 1 RAARTGCCRAARTCDATRTGRGA 23
 DB 6232 AACTGCAAAATCATCATGTGAAA 6254

RESULT 2
 180200/c
 1D 180200 standard; cDNA; 6831 BP.
 AC 180200;
 DT 22-JUN-1998 (first entry)
 DE Phosphatidyl inositol 3-kinase cdk cDNA.
 KM Phosphatidyl inositol 3-kinase; signal transduction; cell cycle;
 KW antagonist; inflammatory joint disease; cell proliferation; cancer;
 KW psoriasis; restenosis; atherosclerosis; therapy; diagnosis; cpk; ss.
 OS Drosophila melanogaster.
 FI Key Location/Qualifiers
 FT 148..5778
 FT /*tag= a
 PN WO9731650-A1.
 PD 04-SEP-1997.
 PF 12-FEB-1997; U02193.
 PR 29-FEB-1996; US-609049.
 PA (REGC) UNIV CALIFORNIA.
 PI Chen Y, Moiz L, Williams LT;
 PI WPI: 97-448442/41.
 DR P-PsDB; W38757.
 DR New isolated phosphatidyl inositol-3 kinase polypeptide - used to
 PT develop products for diagnosis and therapy, particularly for
 PT proliferative disorders, e.g. inflammatory joint diseases, or cancer
 PS Example 1; Fig 9: 77pp; English.
 CC This cDNA sequence codes for cpk (see W38757), a Drosophila
 CC polypeptide that belongs to a novel class of phosphatidyl inositol
 CC 3-kinases that contain a C2 domain, are capable of phosphorylating
 CC a D3 hydroxyl of an inositol ring in phosphatidyl inositol and
 CC phosphatidyl inositol 4-phosphate, but not in phosphatidyl inositol
 CC 4,5-bisphosphate, and which are involved in cell signalling cascades
 CC that control e.g. cell cycle progression and intracellular protein
 CC sorting. Short fragments of cpk cDNA were obtained from a
 CC Drosophila cDNA library by PCR (see T80195-96). These short
 CC fragments were used to screen the cDNA library to obtain larger
 CC fragments, and missing 5' ends were obtained by RACE. A
 CC recombinant host cell, transfected with a vector comprising a cpk
 CC nucleic acid is claimed. Novel phosphatidyl inositol 3-kinases
 CC can be used to screen for agonists/antagonists of activity and
 CC in a claimed method of treating a disorder caused by dysregulation
 CC of a growth factor activation signalling cascade. Antagonists
 CC may reduce Ras activation allowing treatment of proliferative
 CC disorders such as atherosclerosis, inflammatory joint disease,
 CC psoriasis, restenosis following angioplasty, and cancer. 1648 T;
 SQ Sequence 6831 BP; 1829 A; 1685 C; 1669 G; 1648 T;

Query Match 73.9%; Score 17; DB 1; Length 6831;
 Best Local Similarity 65.2%; Pred. No. 24;
 Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

OY 1 RAARTGCCRAARTCDATRTGRGA 23
 DB 4529 AACTGCAAAATCATCATGTGAAA 4507

RESULT 3
 059013/c
 ID 059013 standard; DNA; 38 BP.
 AC 059013;
 DT 12-APR-1994 (first entry)
 DE Primer for p110 cDNA.
 KM Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ss; amplification.
 OS Synthetic.
 PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1; Page 38; 146pp; English.
 CC An SGBAF-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with psvneo. Total RNA was
 CC isolated from the SGBAF-1 and a cDNA library constructed. The
 CC unamplified library was plated on E. coli K12 PLF-F' and screened
 CC with the labelled primer shown and the primer of 059012. Hybridising
 CC clones were sequenced.
 CC See also 051155-6, 059012-23 and 057522-3.
 SQ Sequence 38 BP; 7 A; 6 C; 8 G; 11 T;

Query Match 72.2%; Score 16.6; DB 1; Length 38;
 Best Local Similarity 73.9%; Pred. No. 15;
 Matches 17; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 RAARTGCCRAARTCDATRTGRGA 23
 DB 38 AARTGCCAAARTCATRTGRGA 16

RESULT 4
 057019/c
 ID 057019 standard; DNA; 38 BP.
 AC 057019;
 DT 31-AUG-1994 (first entry)
 DE p110s 3-kinase probe #2.
 KW 110 kd catalytic subunit; phosphatidyl inositol 3-kinase;
 KW transformation; Schistosomaccharomyces pombe; nmt promoter; thiamine;
 KW PtdIns 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Synthetic.
 PN WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 PI WPI: 94-065697/08.
 DR Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Example 1; Page 14; 71pp; English.
 CC The sequences given in 057018-19 are probes which were used in the
 CC isolation of fusion sequences comprising the phosphatidyl inositol
 CC (PtdIns) 3-kinase under the regulatory control of the nmt promoter.
 CC These sequences were transformed into Schistosomaccharomyces pombe cells

CC Immune response against all B. pertussis strains esp. B. parapertussis,
CC the second most important causative agent of whooping cough. The
CC vaccine may also be used to protect against B. bronchiseptica
CC infections.
SS Sequence 1131 BP; 212 A; 365 C; 359 G; 195 T;
Query Match 69.6%; Score 16; DB 1; Length 1131;
Best Local Similarity 63.6%; Pred. No. 54;
Matches 14; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
OY 1 RAARTGCCRAARTCDATRTGAA 22
DB 440 GAACGTCCGAATCGATATGAA 419
RESULT 7
V75788/c
ID V75788 standard; DNA; 641 BP.
AC V75788;
DE 16-MAR-1999 (first entry)
DE Staphylococcus aureus contlg SEQ ID #1477.
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 541..600
FT /tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
EP-786519-A2.
PD 30-JUL-1997.
PD 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMAN-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
PI WPI: 97-374922/35.
DR Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT scored on computer readable medium and used in the production of
PT anti-S. aureus vaccines
PS Claim 1; Page 1969-1970; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S. aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S. aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S. aureus in a sample. S. aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S. aureus DNA sequences contained on the
CC computer readable medium.
SS Sequence 641 BP; 204 A; 95 C; 70 G; 210 T;
Query Match 69.6%; Score 16; DB 1; Length 641;
Best Local Similarity 63.6%; Pred. No. 49;
Matches 14; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
OY 2 AARTGCCRAARTCDATRTGAA 23

DB 238 AATGCCAAATCGATATGAA 217
RESULT 8
ID T80043/c
AC T80043 standard; DNA; 40875 BP.
DE 03-NOV-1997 (first entry)
DE Insert from cosmid 109.
KW Pectinase; yeast; polygalacturonase; phytopathogen; vegetable spoilage;
KW fruit spoilage; food processing; fruit juice; cosmid; ss.
OS Saccharomyces cerevisiae.
FH Key Location/Qualifiers
FT cds 512..1943
FT /tag= a
FT /note= "indicated in the specification as an open reading
FT frame, encoding a protein with homology to
FT S. cerevisiae allantoinase permease (DAL5)"
FT 3622..7107
FT /tag= b
FT /note= "indicated in the specification as an open reading
FT frame"
FT 3965..5188
FT /tag= c
FT /note= "indicated in the specification as an open reading
FT frame"
FT 3461..4546
FT /tag= d
FT /note= "indicated in the specification as an open reading
FT frame, encoding a protein with homologies to
FT pectinases"
FT 5151..6194
FT /tag= e
FT /note= "indicated in the specification as an open reading
FT frame, with homology to S. pombe thiamine
FT repressed gene"
FT 5486..5845
FT /tag= f
FT /note= "indicated in the specification as an open reading
FT frame"
FT 6386..7414
FT /tag= g
FT /note= "indicated in the specification as an open reading
FT frame"
FT 6430..7470
FT /tag= h
FT /note= "indicated in the specification as an open reading
FT frame"
FT 8051..8917
FT /tag= i
FT /note= "indicated in the specification as an open reading
FT frame"
FT 9656..10552
FT /tag= j
FT /note= "indicated in the specification as an open reading
FT frame"
FT 9996..10364
FT /tag= k
FT /note= "indicated in the specification as an open reading
FT frame"
FT 10127..10936
FT /tag= l
FT /note= "indicated in the specification as an open reading
FT frame, encoding a protein with homology to the
FT S. cerevisiae mitochondrial genome maintenance
FT protein (MGM101)"
FT 11161..11523
FT /tag= m
FT /note= "indicated in the specification as an open reading
FT frame"
FT 13441..14517
FT /tag= n
FT /note= "indicated in the specification as an open reading
FT frame"

Query Match	Best Local Similarity	Score	DB 1:	Length	40875:
Matches: 15; Conservative	65.2%	Pred. No. 1.2e+02;			
	5;	Mismatches	3;	Indels	0; Gaps 0

Oy	1	RAARTGCCRAARTCDARTRGAA	23
		: : : : : :	
Db	422	AAAGTCCAAATCAAGCTGCA	400
RESULT	9		
ID	029627/c		
AC	029627	standard; DNA; 2540 BP.	
DT	16-MAR-1993	(first entry)	
DE	Hepatitis C virus HC-J1 5' region.		
ND	Non-A non-B hepatitis; NANBH; HCV; detection; diagnosis; screening;		
KM	PCR; primer: polymerase chain reaction; ss.		
OS	Hepatitis C virus.		
PN	EP-510952-A.		
PD	28-OCT-1992.		
PF	23-APR-1992; 303625.		
PR	26-APR-1991; JP-191376.		
PA	(IMMO.) IMMUNO JAPAN INC.		
PI	Nakamura T, Okamoto H;		
PT	WPI; 92-359137/44.		
PT	Detection of non-A, non-B hepatitis virus - using new		
PT	oligo-nucleotide primers with nucleotide sequences corresp. to		
PS	part. of the viral RNA		
PS	Disclosure; Page 16; 54pp; English.		
CC	This sequence represents the 5' region of hepatitis C virus RNA. The		
CC	original sample was obtained from human and chimpanzee plasma. RNA		
CC	was isolated from several samples and homology compared, and the		
CC	respective sequence of about 1900 - 2500 nucleotides of the 5'		
CC	terminus and 1100 nucleotides of the 3' terminus determined. The 5'		
CC	region (given) contains a non-coding region of at least 340		
CC	nucleotides and a region coding for the structural protein followed		
CC	by a region coding for the non-structural protein (none actually		
CC	detailed on the sequence given in the specification). When compared		
CC	with the sequence of HCV disclosed in EP-388232 this sequence showed		
CC	homology of 95.2%.		
SQ	Sequence	2540 BP; 477 A; 777 C; 737 G; 549 T;	
Query Match		67.8%; Score 15.6; DB 1; Length 2540;	
Best Local Similarity		65.0%; Pred. No. 97;	
Matches	13; Conservative	6; Mismatches	1; Indels 0; Gaps 0;
Oy	3	ARTGCCRAARTCDARTRGAA	22
		: : : : : :	
Db	2193	AAATGCCAAAGCCTATATGGA	2174
RESULT	10		
ID	043888/c		
AC	043888	standard; cDNA to RNA; 2540 BP.	
DT	21-OCT-1993	(first entry)	
DE	NANB hepatitis virus polynucleotide N-2540-1.		
ND	Non-A, non-B; virus; polymerase chain reaction; detection;		
KM	sensitive; specific; HCV; NANBH; ss.		
OS	Non-A, non-B hepatitis virus.		
PN	J05091884-A.		
PD	16-APR-1993.		
PF	10-APR-1991; 196175.		
PR	12-JUN-1990; JP-153401.		
PR	08-NOV-1990; JP-304405.		
PA	(NAKA.) NAKAMURA T.		
DR	WPI; 93-196637/25.		
DR	P-DSDB; R38278.		
PT	Antigen related to non-A and non-B hepatitis virus - comprises		
PT	non-translation region comprising 340 - 341 mols. of nucleotides,		
PT	non-translation region comprising 1885 - 2551 mols. of		
PT	nucleotides including region 1,149 and, etc.		
BS	Claim 2; Page 17-18; 73pp; Japanese.		
CC	The sequence is that of NANB hepatitis virus polynucleotide N-2540-1		
CC	which codes for a non-A, non-B (NANB) hepatitis virus gene HC-OM.		

CC The polypeptide it encodes may be used in a system for detecting
CC NANB hepatitis. This method is highly specific and sensitive, and
CC can detect NANB hepatitis virus which could not be detected by
CC conventional methods.
SQ Sequence 2540 BP; 477 A; 773 C; 741 G; 549 T;

Query Match 67.8%; Score 15.6; DB 1; Length 2540;
Best Local Similarity 65.0%; Pred. No. 97;
Matches 13; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 3 ARTGCCAARTCDATRTGAA 22
1:|||||:|:|:|:|:|:|
DB 2193 ARTGCCAAGCCATATGGA 2174

RESULT 11
ID 048539 standard; cDNA: 2401 BP.

AC 048539;
DT 21-FEB-1994 (first entry)
DE Bacillus subtilis inositol dehydrogenase gene.
KW Recombinant protein production; inositol dehydrogenase; ss.
OS Bacillus subtilis.

FH Key Location/Qualifiers
FT cds 1329..2363
/tag= a
/product= inositol_dehydrogenase

PN J05192163-A.
PD 03-AUG-1993.
PF 30-SEP-1991: 252073.
PR 30-SEP-1991: JP-252073.
PA (SUNR) SUNTOREY LTD.
DR WPI: 93-277476/35.
P-PSDB: R40918.

PT Inositol dehydrogenase gene - Isolated from e.g. bacteria
PI belonging to Bacillus subtilis, in large amt.

PS Claim 4: Page 8-10; 18pp; Japanese.

CC The Bacillus subtilis inositol dehydrogenase gene has been cloned
CC and sequenced. Microorganisms transformed by the coding sequence can
CC be cultured to produce large amounts of recombinant inositol
CC dehydrogenase.
SQ Sequence 2401 BP; 657 A; 517 C; 609 G; 618 T;

Query Match 67.8%; Score 15.6; DB 1; Length 2401;
Best Local Similarity 65.0%; Pred. No. 96;
Matches 13; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 3 ARTGCCAARTCDATRTGAA 22
1:|||||:|:|:|:|:|:|
DB 1998 ARTGCTGAATCATATGGA 2017

RESULT 12
ID 057522/c
AC 057522 standard; cDNA: 381 BP.

DT 12-APR-1994 (first entry)

DE Human P1TR-c cDNA.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KW ss.

PN W09321328-A.

PD 28-OCT-1993.

PF 13-APR-1993: G00761.

PR 13-APR-1993: GB-008135.

PA (LUDW) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI ParkerPJ, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

P-PSDB: R46552.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation

PS Disclosure: Fig 20; 146pp; English.

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC glucose levels can be controlled using the kinase.

CC See also 051155-6, 059012-23 and 057523.

SQ Sequence 381 BP; 116 A; 78 C; 85 G; 102 T;

Query Match 67.8%; Score 15.6; DB 1; Length 381;
Best Local Similarity 72.2%; Pred. No. 70;
Matches 13; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 6 GCCRAARTCDATRTGAA 23
1:|||||:|:|:|:|:|:|
DB 378 GCCGAATCGATGTGGA 361

RESULT 13

ID 057523/c

AC 057523 standard; cDNA: 393 BP.

DT 12-APR-1994 (first entry)

DE Human P1TR-f cDNA.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KW ss.

PN W09321328-A.

PD 28-OCT-1993.

PF 13-APR-1993: G00761.

PR 13-APR-1993: GB-008135.

PA (LUDW) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI ParkerPJ, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

P-PSDB: R46553.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PT activity, useful for controlling cell proliferation

PS Disclosure: Fig 21; 146pp; English.

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC glucose levels can be controlled using the kinase.

CC See also 051155-6, 059012-23 and 057522.

SQ Sequence 393 BP; 87 A; 103 C; 124 G; 79 T;

Query Match 67.8%; Score 15.6; DB 1; Length 393;
Best Local Similarity 72.2%; Pred. No. 70;
Matches 13; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 6 GCCRAARTCDATRTGAA 23

Db 393 GCCGAGTCGATGTGGA 376

RESULT 14

ID 059023 standard; DNA; 24 BP.

AC 059023; 12-APR-1994 (first entry)
 DE Degenerate primer for p110 cDNA.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ss; amplification.
 OS Synthetic.
 FH Key
 FT misc_difference 7 Location/Qualifiers
 FT /tag= a
 FT /note= "inosine"

PD 28-OCT-1993.
 PE 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 2, Page 53; 146pp; English.
 CC Two novel cDNAs related to p110 have been cloned. Degenerate
 CC primers were designed to conserved sequences between human p110 and the
 CC related yeast gene VPS34. These were used in RT-PCR using mRNA from the
 CC human cell lines MOLT4 and U937. Two novel cDNAs, P1TR-C and P1TR-F,
 CC related to p110, were isolated.
 CC See also Q51155-6, Q59012-22 and Q57522-3.
 SQ Sequence 24 BP; 7 A; 3 C; 2 G; 4 T;

Query Match 67.8%; Score 15.6; DB 1; Length 24;
 Best Local Similarity 78.3%; Pred. No. 43;
 Matches 18; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RAATGCCRAARTCDATRTGGA 23

Db 2 AATGCCRAARTCDATRTGGA 24

RESULT 15

ID 063752/c standard; cDNA to genomic RNA; 2540 BP.

AC 063752;
 DE 30-JAN-1995 (first entry)
 DE NANBHV genomic fragment.
 KW Polymerase chain reaction; PCR; primer; amplify; detection; NANBHV;
 KW non-A, non-B hepatitis virus; 5'-terminal region; core protein; ss.
 OS Synthetic.
 PN J06125777-A.
 PD 10-MAY-1994.
 PE 20-JUN-1991; 247120.
 PF 20-JUN-1991; JP-247120.
 PA (NAKA/) NAKAMURA T.
 DR WPI: 94-187937/23.
 PT Oligonucleotide primer pairs specific for non-A, non-B hepatitis
 PT virus - used for high sensitivity detection of non-A non-B (NANB)
 PT hepatitis virus
 PS Disclosure; Page 22-23; 25pp; Japanese.
 CC The sequences given in Q63752-53 represent fragments of the non-A,
 CC non-B hepatitis virus (NANBHV) genome. These fragments were amplified
 CC using the primers given in Q63732-51. These primers were used in the
 CC detection of NANBH. The primers are based on the 5'-terminal region and
 CC the core protein coding region. The method allows highly sensitive
 CC detection of NANBH.

SQ Sequence 2540 BP; 477 A; 777 C; 737 G; 549 T;

Query Match 67.8%; Score 15.6; DB 1; Length 2540;
 Best Local Similarity 65.0%; Pred. No. 97;
 Matches 13; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 3 AATGCCRAARTCDATRTGGA 22

Db 2193 AATGCCAAGCCTATATGGA 2174

Search completed: September 13, 1999, 15:59:38
 Job time: 4805 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:57 ; Search time 1694.61 Seconds

(without alignments)
26.772 Million cell updates/sec

Title: us-09-325-095-29

Perfect score: 23
Sequence: 1 RAAATGCCRAATCDAITRTGAA 23

Scoring table: IDENTITY_NUC

Searched: 2546578 segs, 98626752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17.2	74.8	740	45	AU004743	AU004743 AU004743
2	17	73.9	435	27	M99426	M99426 TGEEST201b1
3	17	73.9	481	29	AA153179	AA153179 m927a09.t
4	17	73.9	269	30	AA240189	AA240189 m927c04.t
5	17	73.9	647	51	A1728321	A1728321 BNGH1104
6	16.6	72.2	376	35	C45088	C45088 C45088 yu1
7	16.6	72.2	267	36	C67578	C67578 C67578 yu1
8	16.6	72.2	756	45	AU002985	AU002985 AU002985
9	16.6	72.2	658	45	AU006090	AU006090 AU006090
10	16.6	72.2	423	53	HSK005641	HSK005641
11	16.4	71.3	373	36	C69479	C69479 C69479 yu1
12	16.2	70.4	319	20	T02063	T02063 WEST02784 E
13	16.2	70.4	362	20	D21659	D21659 HSDR3212 S
14	16.2	70.4	283	23	D60457	D60457 HUM110E12A
15	16.2	70.4	360	36	C61779	C61779 C61779 yu1
16	16.2	70.4	360	36	C67000	C67000 C67000 yu1
17	16.2	70.4	657	46	A1455444	A1455444 ID24279.3
18	16.2	70.4	299	50	AV045367	AV045367 AV045367
19	16	69.6	192	20	Z40168	Z40168 HSC1UD042 n
20	16	69.6	479	25	N76563	N76563 YV40803.r1
21	16	69.6	498	28	AA073214	AA073214 m993c03.r
22	16	69.6	360	35	C42845	C42845 C42845 yu1
23	16	69.6	440	36	AA629895	AA629895 ad45b11.s
24	16	69.6	285	40	AA912476	AA912476 O196e07.s
25	16	69.6	330	40	AA912872	AA912872 O127a03.s
26	16	69.6	224	40	AA922454	AA922454 om63b11.s
27	15.8	68.7	482	27	AA001587	AA001587 z649f01.s
28	15.8	68.7	420	29	AA191170	AA191170 z645a02.r
29	15.8	68.7	528	35	C23190	C23190 C23190 Japa
30	15.8	68.7	300	35	C53168	C53168 C53168 yu1
31	15.8	68.7	600	36	AA601777	AA601777 NCP4E37
32	15.8	68.7	303	38	AA784401	AA784401 d5d08a1.r
33	15.8	68.7	599	38	AA786082	AA786082 j5b0281.r
34	15.8	68.7	526	38	AA786117	AA786117 j5b0581.r
35	15.8	68.7	538	40	AA933666	AA933666 om59b03.s
36	15.8	68.7	335	40	AA965589	AA965589 k5a12a1.r
37	15.8	68.7	511	40	AA966375	AA966375 w4602a1.r
38	15.8	68.7	479	40	AA966448	AA966448 w5b02a1.r
39	15.8	68.7	411	43	A1209373	A1209373 a0c0a1.r
40	15.8	68.7	428	48	A1572092	A1572092 t73e06.x
41	15.8	68.7	768	49	A1620707	A1620707 lu96a12.x
42	15.8	68.7	588	50	A1675336	A1675336 w98b06.x
43	15.6	67.8	346	20	D24320	D24320 RICR1727A R
44	15.6	67.8	399	20	T10141	T10141 seg1375 b4H
45	15.6	67.8	236	24	N20317	N20317 yx44a11.s1

ALIGNMENTS

RESULT 1
AU004743
LOCUS AU004743 740 bp mRNA
DEFINITION AU004743 Bombyx mori p50(Daizo) Bombyx mori cDNA clone ws20797,
ACCESSION AU004743
NID 94162114
VERSION AU004743.1 GI:4162114

KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
Ditrysia; Bombycoidea; Bombycidae; Bombyx.
REFERENCE
1 (bases 1 to 740)
Mita, K., Moriyasu, M., Shimada, T., Okano, K. and Maeda, S.
Establishment of cDNA database of Bombyx mori
Unpublished (1999)
On Jan 14, 1998 this sequence version replaced gi:1798193.
COMMENT

AUTHORS
TITLE
JOURNAL

FEATURES
source
1..740
/organism="Bombyx mori"
/strain="p50(Daizo)"
/db_xref="taxon:7091"
/map="16"
/clone="ws20797"
/clone_1lb="Bombyx mori p50(Daizo)"
BASE COUNT 225 a 163 c 154 g 198 t
ORIGIN

Query Match 74.8%; Score 17.2; DB 45; Length 740;
Best Local Similarity 63.6%; Pred. No. 1.8e+02;
Matches 14; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 RAARTGCCRAARTCDATRTGAA 22
:::|||||:::|::|::|::|
DB 425 AAGTGCCGAGTCACATGAA 446

RESULT 2
W99426 435 bp mRNA EST 11-SEP-1997
LOCUS Tgzzz01b12.r1 TgME49 Tachyzoite cDNA Toxoplasma gondii cDNA clone
DEFINITION tgz01b12.r1 5', mRNA sequence.
ACCESSION W99426
NID 91435312
VERSION W99426.1 GI:1435312
KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
REFERENCE
1 (bases 1 to 435)
Hehl, A., Manger, I., Marra, M., Sibley, L.D., Ajlaka, J.A.,
Aslett, M.A., Dietrich, N., Dubuque, T., Hillier, L., Kucaba, T.,
Wan, K.L., Waterston, R.H. and Boothroyd, J.
WashU-Merck-Stanford-NIH Toxoplasma EST project
Unpublished (1996)
On Jan 25, 1995 this sequence version replaced gi:637759.
COMMENT

AUTHORS
TITLE
JOURNAL

FEATURES
source
Contact: Marra M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxowatson.wustl.edu
Clones are available from Genome Systems (genome@MO.NET); the
library is available from John Boothroyd
(jbooth@popserver.stanford.edu)
Seq primer: T3
High quality sequence stop: 341.
Location/Qualifiers

source
1..435
/organism="Toxoplasma gondii"
/strain="ME49, clone PDS"
/db_xref="taxon:5811"
/map="10"
/clone="tgzz01b12.r1"
/clone_1lb="TgME49 Tachyzoite cDNA"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Zap; Site-1: EcoRI; Site-2: XhoI;
Toxoplasma parasites were grown in human foreskin
fibroblasts cultures in vitro. The library was
constructed by A. Hehl and I. Manger, Stanford University.
cDNAs were synthesized from polyA mRNAs by oligo d(T)
priming and directionally cloned into the EcoRI and XhoI
sites of the Lambda Zap vector using the Zap-cDNA
synthesis kit (Statagene). Warning: the library contains a
small percentage of human cDNAs derived from the human
host cells."
BASE COUNT 118 a 88 c 104 g 123 t 2 others
ORIGIN

Query Match 73.9%; Score 17; DB 27; Length 435;
Best Local Similarity 65.2%; Pred. No. 2.3e+02;
Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAARTGCCRAARTCDATRTGAA 23
:::|||||:::|::|::|::|
DB 51 GAAGTCCGAGTCGTTGAAA 29

RESULT 3
AA153179 481 bp mRNA EST 11-DEC-1996
LOCUS m31a09.r1 Beddington mouse embryonic region Mus musculus cDNA
DEFINITION clone IMAGE:539512 5', mRNA sequence.
ACCESSION AA153179
NID 91724893
VERSION AA153179.1 GI:1724893
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 481)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMNI Mouse EST Project
Unpublished (1996)
On May 17, 1996 this sequence version replaced gi:1182837.
COMMENT

AUTHORS
TITLE
JOURNAL

FEATURES
source
Contact: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 ET
High quality sequence stop: 393.
Location/Qualifiers
1..481
/organism="Mus musculus"
/strain="C57BL6 x DBA"
/db_xref="taxon:10090"
/map="21q22.1-3"
/clone="IMAGE:539512"

```

/clone_lib="Beddington mouse embryonic region"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="7.5dpc"
/lab_host="DH12S"
/note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
Oligo_1: Site_2: Not; Cloned unidirectionally. Primer:
SalI dr. Gastrulating embryos were collected at 7.5dpc
from C57BL/6 x DBA matings, excluding embryos that had
developed head folds and all extraembryonic tissues.
Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
Referenced in Development 121, 2479-2489 (1995)."

```

	Query Match	Score	DB	Length
Best Local Similarity	73.9%	17	29	461
Matches	65.2%			
Conservative	6			
Mismatches	2			
Indels	0			
Gaps	0			

RESULT	4
AA240189	
LOCUS	
DEFINITION	AA240189 269 bp mRNA EST 12-MAR-1997
ACCESION	U72706.r1 Guaywoodford Beller mouse kidney day 0 Mus musculus cDNA clone IMAGE:656262 5', mRNA sequence.
	AA240189

VERSION	*AA240189.1	GI:1864423
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE
AUTHORS

Euryarjota, Metazoa; Chordata; Craniata; Vertebrata, Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 269)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1398069.

CONTACT: Maria M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through ILMN; contact the
 IMAGE Consortium (infoimage.llnl.gov) for further information.
 MGI:402110
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 268.
 Location/Qualifiers
 1..269

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:656262"
/clone_lib="Guaywoodford B612 mouse kidney day 0"
/tissue_type="kidney"
/dev_stage="newborn (day 0)"
/lab_host="SOLR (kanamycin resistant)"
/mouse_organ="kidney; Vector: plasmid SK-; Site: 1:
ECORI; Site: 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size: 1.0 Kb; Uni-ZAP XR Vector;

```

```

-5' adaptor sequence: 5' GAAATCGGACAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTT 3' Library provided
Lisa Guay-Woodford."
BASE COUNT      66 a      55 c      59 g      89 t
ORIGIN

```

Query Match	73.9%	Score 17:	DB 30:	Length 269:
Best Local Similarity	65.2%	Pred. No. 2.4e+02:		
Matches 15:	Conservative 6:	Mismatches 2:	Indels 0:	Gaps 0:
Qy	1	RAARTGCCRAATCATCTGTA	23	
		..TT	
Db	19	GATTGCCAAAGTCATCTGTA	41	

RESULT	5
LOCUS	A1728321/c
DEFINITION	A1728321 647 bp mRNA EST 11-JUN-1999 BINGSHI10458 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to (M74027) mucin [Homo sapiens], mRNA sequence.

ACCESSION	AI/28321
NID	95047173
VERSION	AI728321.1
KEYWORDS	GI:5047173
SOURCE	EST.
	upland cotton.

ORGANISM
Gossypium hirsutum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliopsida; eudicotyledons; core
eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 647)
Brewitt, M., Matz, E.C., Davy, D.F. and Burr, B.
ESTs from developing cotton fiber
Unpublished (1999)

COMMENT On Jun 5, 1998 this sequence version replaced g1:3188104.

Contact: Ben Burr
Biology Department
Brookhaven National Laboratory
Upton, NY 11973, USA
Tel.: 516-344-3396
Fax: 516-344-3407
Email: burr@bnl.bnl.gov
Seq primer: T3 Primer.

FEATURES	Location/Qualifiers
.source	1. .647

BASE COUNT	ORIGIN
137 a	221 c 112 g 175 t 2 others

Query Match	73.9%	Score 17	DB 51	Length 647
Best Local Similarity	65.2%	Pred. No. 2	2e02	
Matches 15	Conservative 6	Mismatches 2	Indels 0	Gaps 0
QY	1	RAARTGCCAARTCDATRTGAA	23	
		: : : : : : : : :		
DB	24	GAACTGCCGCACTCAATGTGGAA	2	

RESULT	6		
LOCUS	C45088	376 bp	EST
DEFINITION	C45088.Yu1 Kohara unpublished cDNA <i>Caenorhabditis elegans</i> cDNA clone yk376a7 5', mRNA sequence.		
ACCESSION	C45088		


```

VERSION      AU006090.1  GI:4163474
KEYWORDS
SOURCE       domestic silkworm.
ORGANISM     Bombyx mori
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
              Ditrysia; Bombycoidea; Bombycidae; Bombyx.
REFERENCE    1 (bases 1 to 658)
AUTHORS      Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
TITLE        Establishment of cDNA database of Bombyx mori
JOURNAL      Unpublished (1999)
COMMENT      On Jan 17, 1998 this sequence version replaced gi:1900538.

Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitsu@nirs.nirs.go.jp
PROJECT - 'CREST project by JST'.

FEATURES
source       Location/Qualifiers
              1..658
              /organism="Bombyx mori"
              /strain="p50(Daizo)"
              /db_xref="taxon:7091"
              /clone="v40467"
              /clone_lib="Bombyx mori p50(Daizo)"

BASE COUNT   232 a      91 c      100 g      235 t

ORIGIN
Query Match      72.2%; Score 16.6; DB 45; Length 658;
Best Local Similarity 66.7%; Pred. No. 3.4e+02;
Matches 14; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY      3 AATGCCCAATCDATRTGAA 23
        |||||:||||:||||:|
        53 AATGCCCAATCAATATGGA 33

RESULT 10
HSM005641/c
ID      HSM005641 standard; RNA; EST; 423 BP.
AC      AL041165;
XX      AL041165.1
SV      AL041165.1
XX      el399502
NI      el399502
DT      12-MAR-1999 (Rel. 59, Created)
DT      12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX      Homo sapiens mRNA; EST DKFZp434G1816_r1 (from clone
DE      DKFZp434G1816)
XX      EST: expressed sequence tag.
XX      EST: expressed sequence tag.
XX      Homo sapiens (human)
OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC      Primates; Catarrhini; Hominoidea; Homo.
XX      [1]
XX      [1]
RN      1-423
RP      Koehrer K., Beyer A., Mewes W., Gassenhuber J., Wiemann S.;
RA      Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
RL      MIPS, Am Klopperspitz 18a D-82152 Martinsried, GERMANY
XX      CC Clone from S. Wiemann, sequenced by BMFZ within the cDNA
CC      sequencing consortium of the German Genome Project
CC      s1 sequence also available
CC      This clone is available at the RZPD in Berlin
CC      Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

```

```

CC      Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX      Key      Location/Qualifiers
FH      source      1..423
FT      /db_xref="taxon:9606"
FT      /organism="Homo sapiens"
FT      /clone="DKFZp434G1816"
FT      /clone_lib="434 (synonym: htes3). Vector pSport1; host
FT      DH10B; sites NotI + SalI"
FT      /dev_stage="adult"
FT      /tissue_type="testis"
XX      SQ      Sequence 423 BP; 143 A; 74 C; 77 G; 126 T; 3 others;

Query Match      72.2%; Score 16.6; DB 53; Length 423;
Best Local Similarity 60.9%; Pred. No. 3.5e+02;
Matches 14; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY      1 RAATGCCCAATCDATRTGAA 23
        :|||:||||:||||:|
        DB      80 GAATGCCCAATCAATGATAA 58

RESULT 11
C69479      373 bp      mRNA      EST      23-SEP-1997
LOCUS      C69479 Yuiji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk364d8 5', mRNA sequence.
ACCESSION   C69479
NID         92430835
VERSION     C69479.1 GI:2430835
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans.
REFERENCE    Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
              Rhabditiina; Rhabditioidae; Rhabditiidae; Peloderiinae; Caenorhabditis.
              1 (bases 1 to 373)
              Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
              Sano,M., Miyata,A. and Nishigaki,A.
              Expression map of the C.elegans genome
              Unpublished (1996)
              On Sep 12, 1996 this sequence version replaced gi:1294543.

TITLE       JOURNAL
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1294543.

Contact: Yuiji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.

FEATURES
source       Location/Qualifiers
              1..373
              /organism="Caenorhabditis elegans"
              /strain="CB1489 him-8(el489)"
              /db_xref="taxon:6239"
              /clone="yk364d8"
              /clone_lib="Yuiji Kohara unpublished cDNA"
              /note="dev_stage=varied, sex=Hermaphrodite male,
              tissue_type=whole animal"

BASE COUNT   108 a      62 c      101 g      101 t      1 others

ORIGIN
Query Match      71.3%; Score 16.4; DB 36; Length 373;
Best Local Similarity 68.2%; Pred. No. 4.4e+02;
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      2 AATGCCCAATCDATRTGAA 23
        |||||:||||:||||:|
        DB      173 AATGCCCAATCAATGATAA 194

```

RESULT 12
 LOCUS T02063/c 319 bp mRNA EST 10-NOV-1992
 DEFINITION *EST02784 Early embryo, Stragene (cat. #937007) Caenorhabditis elegans cDNA clone CESTY83, mRNA sequence.

ACCESSION T02063
 NID 9278544
 VERSION T02063.1 GI:278544
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 319)
 McCombie, W.R., Kelley, J.M., Aubin, L., Goscochea, M., Fitzgerald, M.G., Wu, A., Adams, M.D., Dubnick, M., Kerlavage, A.R., Venter, J.C. and Fields, C.A.
 Caenorhabditis elegans cDNAs
 Unpublished (1993)

TITLE JOURNAL
 COMMENT Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 Reverse.
 Location/Qualifiers

FEATURES
 source 1..319
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone_lib="Early embryo, Stragene (cat. #937007)"

BASE COUNT 86 a 58 c 66 g 109 t

ORIGIN

Query Match 70.4%; Score 16.2; DB 20; Length 319;
 Best Local Similarity 61.9%; Pred. No. 5.5e+02;
 Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 RAATGCCRAATCDATRTGR 21
 :|||:|||||:|||||:|||||:
 Db 251 AAATGCGAATCATATGTA 231

RESULT 13
 LOCUS 221669 362 bp mRNA EST 20-APR-1993
 DEFINITION HSDH3J3E12 Stragene cDNA library Human heart, cat#936208 Homo sapiens cDNA clone H23E12, mRNA sequence.

ACCESSION 221669
 NID 938542
 VERSION 221669.1 GI:38542
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 362)
 Aulfray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y., Sebastiani, K., Kitchin, C. and Tessier, A.
 IMAGE: molecular integration of the analysis of the human genome and its expression
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 95277534

TITLE JOURNAL
 MEDLINE 95277534
 COMMENT Contact: Genzentrum Muenchen
 Laboratorium fuer molekulare Biologie

Am Kioferstr. 18a, 8033 Martinsried, Germany
 Email: obermayer@vms.biochem.mpg.de.

FEATURES
 source 1..362
 /organism="Homo sapiens"
 /db_xref="GDB:0057168E"
 /db_xref="taxon:9606"
 /clone_lib="Stragene cDNA library Human heart, cat#936208"
 /note="Vector: pBluescript SK(+); Human heart cDNA library. Cloning vector pBluescript SK(+)"

BASE COUNT 88 a 103 c 108 g 63 t

ORIGIN

Query Match 70.4%; Score 16.2; DB 20; Length 362;
 Best Local Similarity 61.9%; Pred. No. 5.5e+02;
 Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 RAATGCCRAATCDATRTGR 21
 :|||:|||||:|||||:|||||:
 Db 111 AAGATGCCAAGCTATGTGG 131

RESULT 14
 LOCUS D60457/c 283 bp mRNA EST 28-AUG-1995
 DEFINITION HUM10E12A Clontech human fetal brain polyA+ mRNA (#6535) Homo sapiens cDNA clone GEN-110E12 3', mRNA sequence.

ACCESSION D60457
 NID 9962096
 VERSION D60457.1 GI:962096
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 283)
 Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuge, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takaichi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S. and Nakamura, Y.
 Fujiwara et al. (1995)
 Unpublished (1995)
 On May 8, 1995 this sequence version replaced gi:801069.

TITLE JOURNAL
 COMMENT Contact: Tsutomu Fujiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035
 Insert Length: 1889 Std Error: 0.00
 High quality sequence stop: 382.
 Location/Qualifiers

FEATURES
 source 1..283
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="22"
 /clone="GEN-110E12"
 /clone_lib="Clontech human fetal brain polyA+ mRNA (#6535)"

BASE COUNT 92 a 50 c 47 g 88 t 6 others

ORIGIN

Query Match 70.4%; Score 16.2; DB 23; Length 283;
 Best Local Similarity 60.9%; Pred. No. 5.6e+02;
 Matches 14; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAATGCCRAATCDATRTGAA 23
 :|||:|||||:|||||:|||||:

Db 277 AATGTGCAAAATCAATGTAGAA 255

RESULT 15

LOCUS C61779

DEFINITION C61779 Yuj1 Kohara unpublished cDNA *Caenorhabditis elegans* cDNA

ACCESSION C61779

VERSION 92420484

KEYWORDS C61779.1 GI:2420484

SOURCE EST.

ORGANISM *Caenorhabditis elegans*.

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiida; Rhabditiina; Rhabditiidae; Rhabditiidae; Peloderiinae; *Caenorhabditis*.

AUTHORS Kohara,Y., Mochizuki,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.

TITLE Expression map of the *C.elegans* genome

JOURNAL Unpublished (1996)

COMMENT On May 9, 1995 this sequence version replaced gi:802443.

CONTACT: Yuj1 Kohara

Gene Library Lab

National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan

Tel: 0559-75-0771

Fax: 0559-75-6240

Email: ykohara@dbj.nig.ac.jp

Location/Qualifiers

1..360

/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"

/db_xref="taxon:6239"

/map="12p"

/clone="YK2488"

/note="dev-stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"

BASE COUNT 110 a 83 c 78 t 1 others

ORIGIN

Query Match 70.4%; Score 16.2; DB 36; Length 360;

Best Local Similarity 61.9%; Pred. No. 5.5e+02;

Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 AATGTGCAAAATCAATGTAG 21

Db 139 AATGTGCAAAATCAATGTAG 159

Search completed: September 13, 1999, 15:45:57

Job time: 4045 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

variation	5551. .5553	/note="clone 90U6; ACG in this entry; substitution"
	5604. .55985	/replace="atg"
repeat_region	6005. .6142	/note="MEM5A repeat: matches 1. .189 of consensus"
repeat_region	6287. .6435	/note="MR repeat: matches 45. .184 of consensus"
repeat_region	7270. .7371	/note="MEM5A repeat: matches 189. .12 of consensus"
variation	7366. .7373	/note="51 copies of 2 mer 83 & conserved"
variation	7473. .7479	/note="clone 90U6; CACACAG in this entry; insertion
	7651. .7654	/replace="cagg"
variation	7851. .7854	/note="clone 90U6; AGAAGT in this entry; insertion
	8607. .8609	/replace="tttcc"
variation	8607. .8609	/replace="tttcc"
repeat_region	8921. .9617	/note="clone 90U6; TAT in this entry; substitution"
variation	9530. .9632	/note="LIME3 repeat: matches 909. .190 of consensus"
variation	10126. .10128	/note="clone 90U6; GTC in this entry; substitution"
repeat_region	10293. .10361	/replace="cgg"
variation	10337. .10339	/note="MR repeat: matches 77. .144 of consensus"
repeat_region	12714. .12790	/note="MR2 repeat: matches 146. .102 of consensus"
repeat_region	12903. .12936	/note="MR2 repeat: matches 47. .124 of consensus"
repeat_region	13166. .13545	/note="U2 repeat: matches 34. .1 of consensus"
variation	13242. .13244	/note="LMB5 repeat: matches 535. .921 of consensus"
variation	13656. .13658	/note="clone 90U6; ACA in this entry; substitution"
	13916. .13918	/replace="ttt"
variation	14086. .14088	/note="clone 90U6; CAA in this entry; substitution"
	14151. .14153	/replace="aaa"
variation	14284. .14286	/note="clone 90U6; AGA in this entry; substitution"
	14609. .14698	/replace="ggc"
repeat_region	15373. .15666	/note="Aly repeat: matches 295. .7 of consensus"
variation	15617. .15617	/note="Aly repeat: matches 293. .1 of consensus"
	15798. .15800	/replace="ctt"
variation	15798. .15800	/note="clone 90U6; AGG in this entry; substitution"

```

repeat_region      /replace="aad"
                  /note="MTrib repeat: matches 390. .1 of consensus"
                  /15814. .16224

Query Match      52.6%; Score 22.6; DB 9; Length 15147;
Best Local Similarity 52.6%; Pred. No.1e+02;
Matches 20; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

Cy      6 YGAYYTRGCGNARGARARATGCCAARFCDDTTRTGRAA 43
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 100513 TGATMAAAGCAGACTATGCCAAGTCATGTGGAA 100550

RESULT      4
LOCUS      SOYPI3K5X      2987 bp      mRNA      PLN      30-DEC-1995
DEFINITION Glycine max phosphatidylinositol 3-kinase mRNA, complete cds.
ACCESSION   L27265
VERSION     L27265.1 GI:736338
KEYWORDS    phosphatidylinositol 3-kinase.
SOURCE      Glycine max (strain Pilze) (tissue library: lambda zap II) root
            module cDNA to mRNA.
ORGANISM    Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1 (bases 1 to 2987)
AUTHORS    Hong,Z. and Yerna,D.P.
TITLE      A phosphatidylinositol 3-kinase is induced during soybean nodule
            organogenesis and is associated with membrane proliferation
            Proc. Natl. Acad. Sci. U.S.A. 91 (20), 9617-9621 (1994)
FEATURES
MEDLINE
JOURNAL
SOURCE      location/Qualifiers
            1. 2987
            /organism="Glycine max"
            /strain="Pilze"
            /db_xref="taxon:3847"
            /tissue_type="root nodule"
            /tissue_lib="lambda zap II"
            62. 2506
            /codon_start=1
            /function="phosphorylation of phosphatidylinositol at D-3
            position of the inositol ring"
            /product="phosphatidylinositol 3-kinase"
            /protein_id="AA83995.1"
            /db_xref="PID:9736339"
            /db_xref="GI:736339"
            /translation="MTGNERFRFLSCDISPVYFVYERLEGNPLNPKSPDLETNAP
            TEKRTRELAYECALYIDGAPFGLPTRRLSSGSPSCWNELITLTKRDLTKQSQT
            FTVWDISHSGGLIGATITLLENKKQLKTKGOKRLWAGKEAGTETPTSPGVPRHE
            RGLERLEKLVNKEYEQIORDVMDLRLPEKMERIKERSLNGSSHLYLVDPFSF
            EHHVYQESGANLFPSPISANDIYVMPPEVGKINPSEHKQLKLSLTREGVIDD
            LKSSSEKRSIORILKYPPTRLSGDEROLMKFRPSLMSKERALTKRELKCVMSDVQ
            EAQALALMGKMGMDIVCALTELISPFSESEYRAVASVLERADDEELQCYLLQVQ
            ALPFESDKRSLSHPLVORALRNITELASPLRWYVAVELYDPAKAPFYCYELIENNM
            MKTAAGNGEGEDGFKQMSLVROTELTAQCSITREVRNKGWTKRTETLRQLSOL
            LSELTYFDEPDIRSLAPAGVLIAGIVSESIFFSALHPLLSRTANGGCKIIFKG
            DDLRQDLQVQWMSLMDRLKLTENLDLHPVLVLTATGQEGEMLEFIPSLAQIISE
            NRSIIISLQKHPDDGEGITATCTLETETFKSAGSVIITYIGIDRHLDNLLND
            GGFHVEFGFELTRDPKPPPMKICKEMAEAGASQSYVTRFKSGCAVYTIKRS
            SNLIINFYIMAGSNNIPDLASDEKGIKNGEFRDLDDERISHRFQDLINSEVAL
            FPGWETIHRMAQYNR"
BASE COUNT      849 a      558 c      658 g      922 t
ORIGIN
polya_site
2987

```



```

repeat_region /rpt_family="MER2_type"
15921..16228
repeat_region /rpt_family="L1"
16381..16626
repeat_region /rpt_family="L2"
18709..18781
repeat_region /rpt_family="L1"
18893..19442
repeat_region /rpt_family="L1"
19581..19797
repeat_region /rpt_family="L1"
19979..20033
repeat_region /rpt_family="MER"
20932..21230
repeat_region /rpt_family="Alu"
22451..22490
repeat_region /rpt_family="MER"
22642..22787
repeat_region /rpt_family="L1"
23587..23789
repeat_region /rpt_family="MER"
25713..25819
repeat_region /rpt_family="MER1_type"
25820..26073
misc_feature /gene="GRB10"
/Note="match to EST AA129408 (NTD:91669294) zn83c02.r1"
26129..26219
misc_feature /gene="GRB10"
/Note="match to EST AA129408 (NTD:91669294) zn83c02.r1"
26640..26937
repeat_region /rpt_family="Alu"
27447..27623
repeat_region /rpt_family="MER53"
28767..28844
repeat_region /rpt_family="MER"
28921..29309
repeat_region /rpt_family="L1"
34010..34297
repeat_region /rpt_family="Alu"
34861..34997
repeat_region /rpt_family="L1"
35458..35758
repeat_region /rpt_family="Alu"
35872..37519
repeat_region /rpt_family="L1"
38300..38850
repeat_region /rpt_family="L1"
39174..39265
repeat_region /rpt_family="MER"
39607..39664
repeat_region /rpt_family="L1"
41280..41542
repeat_region /rpt_family="Alu"
42628..42737
repeat_region /rpt_family="MER"
43326..43518
repeat_region /rpt_family="MER2_type"
44122..44319
repeat_region /rpt_family="MER"
45297..45597
repeat_region /rpt_family="Alu"

```

Query Match 51.6%; Score 22.2; DB 11; Length 124191;
 Best Local Similarity 51.2%; Pred. No. 1.4e+02;
 Matches 22; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

1 GNGAYGAYTRCGNCGARBARATGCCRAARTCDATRTGAA 43

DB 29902 GGAGATGATTGTGGCTGAAATAATTCAAAACGATATTAA 29944

RESULT 6
 OLIVNEWYORK

```

LOCUS OLIVNEWYORK 1999 bp DNA
DEFINITION Ovine lentivirus DNA sequence.
ACCESSION L19198
NID 9309627
VERSION L19198.1 GI:309627
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
FEATURES
source
1. 1999
/organism="Ovine lentivirus"
/isolate="Cornell"
/strain="North American"
/db_xref="taxon:11663"
BASE COUNT 754 a 315 c 512 g 418 t
ORIGIN

Query Match 51.6%; Score 22.2; DB 17; Length 1999;
Best Local Similarity 55.0%; Pred. No. 86;
Matches 22; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 4 GAYGAYTRCGNCGARBARATGCCRAARTCDATRTGAA 43
DB 1215 GAAGAAAGCAGTAAGAGAAATGCAATGATATGAA 1254

RESULT 7
AC007089 118431 bp DNA HTG 16-MAR-1999
LOCUS Homo sapiens clone Dj0616M23, WORKING DRAFT SEQUENCE, 16 unordered
pieces.
ACCESSION AC007089
NID 94417340
VERSION AC007089.1 GI:4417340
KEYWORDS HTG; HTGS_Phasel.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 118431)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1972: contig of 1972 bp in length
* 1973 1990: gap of unknown length
* 1991 3960: contig of 1970 bp in length
* 3961 3978: gap of unknown length
* 3979 7474: contig of 3496 bp in length
* 7475 7492: gap of unknown length

```



```
FEATURES
Source
Location/Qualifiers
1..1138
/organism="Dichelobacter nodosus"
/insertion_seq="IS1253"
/db_xref="taxon:870"
/note="A198; compl associated copy of IS1253"
186..211
/note="nix1; cross-over site"
/organism="Dichelobacter nodosus"
/evidence=experimental
199..2043
/gene="ompB"
<199..2043
/gene="ompB"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="OmpB"
/protein_id="AAB12360.1"
/db_xref="PID:g407292"
/db_xref="GI:407292"
/transl_table="TLSKIDGLEDAYKKKKOLENONPADKVVDPYDPTQGVGNL
GKASOGTTOYVREGSKPDRVNNKEKIASATSYGLVDNONPDLNIVYAKVND
GKVLIOYIGYERAEIKDPKSLQNMKNFNATKIDGYNLAKDLPAKIKGVDOYKIAI
ASYESLSDTINPDQEKYLSKSLQANLEASASKESEYLVDPKHDGSRGTGANYS
AQEFVLEGDAAQTQAVLTADGVSVTEVAAPYDNGVATSTTGAASGLVGN
FKQETFEAAIOGTDEESOSTRIPGRQYNDADQISKENYAGAKVIPSDFEISNL
FTGKGVKTEVDKHDNLIKAVGDLGADINKGTGVYHLGVKPKLEHVQYRVYTNL
DPKLTADDEOKSADIKKRAAFNMKNADKESLADNEFEYRGTKATLEOKKALPKDRT
LVYQHALMGTGIONSFHGKSDNDVPSPADAPNYSALIGNFVATLADICKEVYTGIV
YNWAKTEKTEPPDKLYKEEGYFGYTVJAKAERTYKDKTNDADFRASFTGERASAM
GGSNSVTDEKTSISDDVGVGFAGKLAEDPIKEDDENSEFLGDKPKP"
/gene="ompE"
complement(2115..4031)
/gene="ompE"
complement(2115..>4031)
/gene="ompE"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="OmpE"
/protein_id="AAB12361.1"
/db_xref="PID:g407293"
/db_xref="GI:407293"
/transl_table="TLRQIAPELAAKTEAEKKRGANLDKVVDPYDPTQGVVDGDL
ASVAGQYVREGSKNEPKQNPENSAPTITGVGLDQNPKLNIFFAKSDDKDK
LVAIOYVGPDKNDVDLIDGSDSQONKNVADRLALSHGQPSDELEYANGQIO
LQPSGAVLMLQAOILIDVNVDTQITIKKALASYEKLEQDKREARVAKKALIDKAL
ADLEKIKADKAKYERTMLVLDKYSKITIPMGDYIPGSKILVGGSGAKALVREVEGT
GGVAGPLIAYNFRADEIRAVAEYKISGATTTLESNIOGTBENSYSSTRITFGKNSD
GIDVANSYTGATTVSEASLAERGATAGSTGLVGFADNKAATFDDKGFEEFYKGDSS
NKVQANRHSIAAEAYILENVQYGRVGDIDAEALAAKPEDEIKHIOQPYTRND
KAVOYFYRGCTNATTLKEMALPTDKILVYOGHALMGTGIONSFHGKNSLSYAFGG
KDMVTAIGNFRAEVDLNKDKVAGVAVYVMQGETSELIPDLVAFEGNVAGNTVYGA
ERTYKTEADPRASFEGKEANEMGGSFNSMLADGTYHEDGVAGYGAQRTYEPYK
ESFYDNGMLISHG"
4019..4044
/note="nix2; cross-over site"
/organism="Dichelobacter nodosus"
/evidence=experimental
4332..6460
/gene="ompD"
4332..6460
/gene="ompD"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="OmpD"
/protein_id="AAB12362.1"
/db_xref="PID:g407294"
/db_xref="GI:407294"
/transl_table="MKLSHLAIIISAITLAAGCGNDKPMPODPNGALVBOBNA
IKKATAYLEDKDGKAPALTKDEANTLSQIAPELAAKTEAEKKRGANPADVYDPT
```

```
QTVGVDPFPEKQSSYGVGYLRGKASDPTDQTPVSEHNSAPTSTVGEGLDEQNEL
TNLFPAKTDKTKGLIAIOFVGIKPITDVGDITNNDSIQOONKHAVALDLSLSTG
NENPELOANGVNNKRTKNFWGAGNNAATVKNENIENEIEALIKATIEFDGILYAK
ADPTEALKNKALILEKTLADLOAIQKNKQYERETDLVLDKYSKITIPMGSDIKPTDITQ
KDAITVDVDSASAGAPGLIAYNFAKERTAEVATLVDKESKITDPERFGEVGAIOGTEEN
SVSSTRIPGKNSPDKIKDANSYKATITVSAASIPKAGADPTDITLFPDARKGTL
GVNDATTFNNAHVAGADTQNTIRSLAETNEFDHVQYGRYTGILEKELEAPPEAV
THIODATFNDKSAVNOYFYRGCTATTELOMAALPKDELHLYOGHALMGTGIONSFH
GKNSDLPYAFAGSNTAIGNFVADKDLNKGSEVGYVYVMQGETSELISDKLVDPK
GVNCHVIGAKRTYSTDKNPADFRASFEGQAGEMGGSFNSMLADGTYHEDGV
WGVFGAORTVEPVKKSEXTDMLISHG"
4332..4375
/gene="ompD"
/evidence=experimental
/product="lipoprotein signal peptide"
4516..4542
/gene="ompD"
/note="nix3; cross-over site"
/organism="Dichelobacter nodosus"
/evidence=experimental
complement(6534..8357)
/gene="ompA"
complement(6534..>8357)
/gene="ompA"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="OmpA"
/protein_id="AAB12363.1"
/db_xref="PID:g407295"
/db_xref="GI:407295"
/transl_table="TUSQIDENLFNEKKKQKLENOASDDEVDPYTKGVGNL
OAGQGTQYVREGKSFDAENNEKESATALLGVAKTONPWLNTVILARVADGT
VSIQYIAGNDSDAADFTNVKSLQANDNDNVKFSDLKTNKPLDEAKLEAKROVDLI
IKYERKIDSEITTEKEELEKFEVALKEELNADIDISPEGYLDEKAKLSEYAT
GSDDIIFGDDVPNNSVLGTADGAGCANPVPIYSGRGADATGAGALVGNMFK
VCGTOAIOGTEKSSSTFRGQRYNTQSRAPAKNSYADAAVIALSSEKLDANNAIA
GITDQSELNATIDMKNAKINHLAVKPKLDIVQYGRYTNLPIKVSLETDDEKAK
LENGERTTRAAFAARDPKGERADNFEYRGIDATTIEQMAALPTDKILNYOGHALM
GIDNSFHGAKKSQFSTAGTAGALHAIGNFSAEVDLSKGVGGHYVNWAGSKDRV
IDSLVFRGVHGNVTIVGTAERTYKAGDADFRASFPEDQAGEMGGSFNSVTTAEKYSS
DVWGVGFGRKRTVEPVKSDDNGFLGDDLPDTTKPKPKP"
8345..8370
/note="nix4; cross-over site"
/organism="Dichelobacter nodosus"
/evidence=experimental
8664..10361
/note="IS1253"
complement(8684..8722)
/note="terminal"
complement(8735..9088)
/gene="ORF117"
complement(8735..9088)
/gene="ORF117"
/codon_start=1
/transl_table=11
/product="ORF117"
/protein_id="AAB12364.1"
/db_xref="PID:g1488248"
/db_xref="GI:1488248"
/transl_table="MLVTKYRKCFISANLIDLKELYTDLCRKEVEYLEPFGEADHY
HLLMDYFPMIPSKFTNNLKTVTSRLRMFEAHLAEFYKRWVLMTRAYCLLTGGAT
IDIRQYIEKQERP"
9089..9099
/tpc_type=direct
9100..9110
/tpc_type=direct
9110..10237
/gene="ORF375"
9110..10237
/gene="ORF375"
/codon_start=1
/transl_table=11
```



```

gene
985..4254
/gstraln="Bristol N2"
/gene="C4569.6"
join(985..1062,1269..1451,1635..2019,2534..2661,
2710..2827,2932..3120,3173..3417,3656..3781,3865..4254)
/gene="C4569.6"
/codon_start=1
/evidence-not_experimental
/protein_id="AAA62549.1"
/db_xref="PID:9687871"
/db_xref="GI:687871"
/translation="MTALLLICSTITOPSDGVVIGGQGVFOIRANPLSFSSCD
HLLVTFESNNHRCGTHTYDCTTIFCKKLCCLMKGKPTIDPGNONNPSPTNG
GIGGVYFGGGNGSVQDEEGVCSNNDCKSLCTASVGVKICLSTNGGGNG
PFSNGGCGTSSNCQGVYVNGNQIQTGVSFAPRGWVAPSSNITVDEN
EDVFSKNAPEGRKINSGCERADCDDELSCYTFGEMCRSPKPLPLRCSDAECF
STEYLCVFSTADQRCYKYGVYVYIPIKHKISMELKSTTSSPTITHLAK
PESNGLFAESEAFEPGSLTSLALOKRADMSONGPTPMVVTMSDIPSHVITAG
HGEAVRITKYVKEKEENENGEREYKRIPIKBSGVGVDEPIDPATCKGSCS
GVYRVDNRVYCRIDMEKHKRCCLHSDCISGRCPTGRDIACTEDTISATISIQ
CFYDECSGCEKTLVDEKERKFCRPSPTSPRANOICTTNSOCFQOVCAEKNNAE
LKITISYIPESNFTONSISYGCFSOMSKLSVTFEESVSVI"
complement(10554..11408)
/gene="C4569.7"
complement(join(10554..10673,11154..11408))
/gene="C4569.7"
/note="similar to tight junction protein (ZO-1)
(SP:201_HUMAN, Q07157); coded for by C. elegans cDNA
CESE55F"
/codon_start=1
/protein_id="AAA62550.1"
/db_xref="PID:9687872"
/db_xref="GI:687872"
/translation="MTAYGMPEGEALIECLSLAVELHKOEVIDAHQVYIRNGFKIGG
IDQDPAKPFKIPDSGVYITNVEGSPADVAGLRKHDKILVNGADFTMTDRAYKE
IKOSVYLMHLVLRADLPVSLH"
12390..14157
/gene="C4569.5"
join(12390..12512,12560..12679,12728..12827,12875..12959,
13551..13688,13773..14157)
/gene="C4569.5"
/codon_start=1
/evidence-not_experimental
/protein_id="AAA62548.1"
/db_xref="PID:9687870"
/db_xref="GI:687870"
/translation="MTOEPSDTSYQREDEMLLHPREGARVYVNVYDQNNIVAS
KGLYFNGKTLRTLHEDKKFKMAPFOVSYKDRITTSIAEFSDDRTILRVSLD
GVVHESFEGLLNMTITDGDWYIVKQAEHGNMFMTHLDTPIGMQDVAVESKP
GEAFSRICALDKRTIVASINDPMMNESQRIYFDLENTTGSTSRAGKPEGTEF
PFIKRYGDEFTLNDKSGFSEFKNGEVALRAKIDALVSGFVKGGEALMALVTGM
VADPNEQLCDWMLLETKLDGSSKRAEDRKRQEDBOANK"
complement(14436..18535)
/gene="C4569.8"
complement(join(14436..14510,14558..14863,14917..15332,
15809..16070,16985..17076,17435..17503,17552..17604,
17689..17755,18478..18535))
/gene="C4569.8"
/codon_start=1
/evidence-not_experimental
/protein_id="AAA62551.1"
/db_xref="PID:9687873"
/db_xref="GI:687873"
/translation="MRLVCGFFSFYFFSRKQTSYFRWVSTETSNNTRSKHKLK
SSIMTSLSRMQIDRDLRSTHEVEIYPOETCVHHSDDVLTETCATMSTHVP
NSSYISKATSKRTGSKKVTYRSKSKRTTITGNSKATVPDNDKXS
KDOORQTRKQDEKWEKCDVTRCKVIOPIKLAENQLRIBEDONITYSR
DIDEADMERICKKPAKKNPITMAEPIDERTKIILDRYTKPKPYLAKNNEGCLIFD
RBSFYKSRKKAKSSNLEEDTYGVQPKIDVORPPQNVFSRPGVPAVALLP
TEEDLVADPSISVGETHEMLHKLQVPQITKSKLILNALQPLSIDEDRDHIFTP
ELVFENITRSLVHAQEMEGKRNESKDEGKSKSLFVIEKOESEFYSRANPIETAR
MARRRNRSGSSSRK"
complement(18908..20020)

```

```

CDS
/gene="C4569.9"
complement(join(18908..19103,19235..19661,19710..19944,
19994..20020))
/gene="C4569.9"
/note="weak similarity to C. tentans Balbiani ring 2.1
(PIR:A45294)"
/codon_start=1
/evidence-not_experimental
/protein_id="AAA62552.1"
/db_xref="PID:9687874"
/db_xref="GI:687874"
/translation="MTTALTPDKKIVSPKPTTSDSKTKPRSSSTSKRRSKK
LFCCAKKRTKRSKSKRTKRSKPAKPAKPAKPAKPAKPAKPAKPAKPAK
QSLVDAQATVTPRPPTPIPTGVAPDEAPREPLQPRSVSSTPTSATGT
EQMVAPATLUPPSSKSLHPDDPGDASSRVQRYAEYKEDQDDQDKLRS
VAYPRKFFNTQYVDECRVRWYEDSVPLMSENMHMLRMAISNRIACQSDKAR
CDMDLMEMLILGNG"
complement(21146..26146)
/gene="C4569.10"
complement(join(21146..21383,21527..21953,22344..22487,
22554..22840,23258..23618,24037..24180,24454..24740,
24830..25419,25464..25591,25745..25906,25956..26048,
26092..26146))
/gene="C4569.10"
/note="coded for by C. elegans cDNA yk28c6.3; coded for by
C. elegans cDNA yk28c6.3; coded for by C. elegans cDNA
yk28c6.5; coded for by C. elegans cDNA yk26e8.5"
/codon_start=1
/protein_id="AAA62553.1"
/db_xref="PID:9687875"
/db_xref="GI:687875"
/translation="WMLNLVFPATSVLLAEGFDCNCLLTDKILIKTSPGNVS
VWLINEVTFGEBSKSLILPPOFSIKLEKRVNTPLVNVLPIRBDYRHNVTY
IVSKEMHSLQFLQSLILKIVDRSGSKDLILYLRAEGQOTSKLILAPNRDP
CORPAMQDTHYSQVONAMRNSHVSGRCTALFLNTGDEPNYVTSKNEIRL
LNTKILNANVAYLVLEFETLKPGKAEVNOVORISKTRFYGVPITVPEIDF
KKSSEFNENPNQISIRKTPLYDPTENALSPDGEVININRSTNSSELYEVE
KTEKHFFPMQSOEGIFRYIRRHDSNTLITIPFDIRNELSEENRLAYCYTMD
FYNVKITTGVQNSNNTVOMKKSSTNSIESVNLIPAREPYVDTVMIIYSEHDGL
LRFDVENMTIEVKTYKTYLRETLVAKYPLINFGDEPYTNTVDIIISNNNSKV
ETASVOINGKSGSYIRIGNPYYDYDFINYPNSSEKWNDESQGVYMFINDPFR
KKSIGKPDENNNDIORSPDDYDIAVIVSNKTKKNDPAITVRSLETGILYEAQY
ENPWFYSIREPGIRNELVEHLSIERPDYVHVITKNHSIGTSENSFNQWEX
QENACTIAEYGVHNTYINGETILUEEKSLIFDPKNAIVLILLSSKCSLDVN
REFANKESYQVLEWCPDSKGYEDITYVKSNNQNRVAYQISRFERPOQTSPT
YIPVTSYPTTNPSPALPNLIDFLNERNENVKYDSTELAMICSEIENGCTTII
TRVLPVNTYITIAKGTGIAFINKSOKTAYRLTEKSIIMGFENSYSYSEHCG
NEWYITSEGEKRNKRESILATLENNSKRTEVNVQVAYKSFDAKFKLSS"
26971..28084
/gene="C4569.4"
join(26971..26997,27047..27281,27337..27577,27889..28084)
/gene="C4569.4"
/note="weakly similar to C. tentans Balbiani ring 2.1
(PIR:A45294)"
/codon_start=1
/evidence-not_experimental
/protein_id="AAA62556.1"
/db_xref="PID:9687878"
/db_xref="GI:687878"
/translation="MTTALTPDKKIVSPKPTTSDSKTKPRSSSTSKRRSKG
LFCCAKKRTKRSKSKRTKRSKPAKPAKPAKPAKPAKPAKPAKPAKPAK
GIVDADAKATVPRPPTPIPTGVAPDEAPREPLQPRSVSSTPTSATGT
YPSHKTFTQVQKDECRVRWYEDSVPLMSENMHMLRMAISNRIACQSDKATCD
MMKLMEETELIDNG"
complement(28258..30680)
/gene="C4569.11"

```

Query Match 50.7%: Score 21.8; DB 36; Length 40293;
Best Local Similarity 53.5%: Pred. No. 1.8e+02;
Matches 23; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTTCGNCARAPARTGCCRAARTCATYRTGRAA 43

DB 7566 GGGGACAAAAGTGTGAGGAAAGTGTCAAAATTGCTGTGAAA 7608

RESULT 14
CS39KBCIV/c
LOCUS CS39KBCIV 39090 bp DNA PLN 07-MAR-1997
DEFINITION S.cerevisiae 39kb DNA segment of chromosome IV.
ACCESSION X99000
X99000.1 GI:1429338
NID 91429338
KEYWORDS anti-silencing protein; ASF2 gene; GABA transporter protein; gdh gene; Gly-tRNA; HEN5 gene; high-mobility-group-like protein; hydroxymethylbilanesynthase; methylated-DNA-protein-cysteine demethyltransferase; MGT1 gene; NAD-dependent glutamate dehydrogenase; NHP2 gene; secretory component; SHM1 gene; SHR3 gene; sugar transport protein; UGA4 gene.
baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 39090)
AUTHORS Bahr, A., Moller-Rieker, S., Hankeln, T., Kraemer, C., Protl, U. and Schmidt, E. R.
TITLE The nucleotide sequence of a 39 kb segment of yeast chromosome IV: 12 new open reading frames, nine known genes and one genes for Gly-tRNA
JOURNAL Yeast 13 (2), 163-169 (1997)
MEDLINE 97197981
REFERENCE 2 (bases 1 to 39090)
AUTHORS Schmidt, E. R.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1996) E. R. Schmidt, Johannes Gutenberg-Universitaet, Institut fuer Molekulargenetik, J.J. Becherweg 32, D-55099, Mainz, FRG

FEATURES
SOURCE Location/Qualifiers
1. 39090
/organism="Saccharomyces cerevisiae"
/strain="FY1679"
/db_xref="taxon:4932"
/chromosome="IV"
/clone="cosmid 1C10"
/map="left arm"
complement(636..1259)
/note="ORF D0884"
/codon_start=1
/protein_id="CA67473.1"
/db_xref="PID:e253394"
/db_xref="PID:91429339"
/db_xref="GI:1429339"
/db_xref="SWISS-PROT:Q12328"
/translation="MYTGFLEQISPAQKKPYNELPEEGERGAEAMINMPTSCPG KSVGVGTGALGVGLGEMASMAVDPLPTPATATAGAGTGGISRTPOOI SDLPPOOMKLOFTDMKRSYSAKNFYICIMYAGVACVIESLARNDIYNGTAGI FTGAGATKACAPQALMGAGFAFSAIDIKNSDGRPNQDKE"
complement(1650..3017)
/note="ORF D0888"
/codon_start=1
/protein_id="CA67474.1"
/db_xref="PID:e253333"
/db_xref="PID:91429340"
/db_xref="GI:1429340"
/db_xref="SPTREMBL:Q12468"
/translation="MANNDRKSIKTAGGSMISNKTVEKLQRLKERYTVEDELSEI ALSWRKPEOEPEHALSOSLSLKTILKOOSSTDIPSTHVLISKISCERTHYAVR AGNIEWGIILGFTLKNIVVDCENLPVGETEPYNAOLSEYMYOYIDEMVAND GGDGRDYKGAALNVGFRSHSPGIDCMISNIDIQODLNORPDYVAIVVDPKLSLE DKILMGAFRTIESKSDNSATSYIELTIIIFDSLNALPEFKLNLMLVCDEDESDQ ISLNRLDSMKOYSLYLMDSKNVRIKILATIESVSNENKNIDYONSTSOFCIANT ORGDSLETSSGMSGDNSTSDVMDENRNLTEFDSTSLKINGEPIHVNREVSRS STDNHNKRRKRNNSQERCHDEGNDMLQRNVLETDIYAAKNRIILASKIKOYERLRYK DTFLL"
gene complement(3293..6571)

CDS
/gene="GDB"
complement(3293..6571)
/gene="GDB"
/note="ORF D0892"
/codon_start=1
/product="NAD-dependent glutamate dehydrogenase"
/protein_id="CA67475.1"
/db_xref="PID:e253395"
/db_xref="PID:91429341"
/db_xref="GI:1429341"
/db_xref="SWISS-PROT:P33327"
/translation="MLDNKRKGLNSLNPDTASLSISSMSDYHYVDPEKQDREE VIDLLDQGEFLPDLLDQEDVDFNLSGLIDLFPSRSPQILSIHSLSKLDLFFR KSKNGIOPRLFSIKNITIDNHAIFMESSTGVISDSQKCNKFAISDAVNDLLEH GKDPIKRNRIEMDSCPPYELDSIDLFLDNKSOQKCNPLVFPAPESKDLFFVYS VYPPDKPDGVOISSODLLKGDISEISDKTWKVSNNENKYLFGALLVYREPPVIR TTRAEVNDQEDRIILVAYKRPRTKRYEASLILFHYVLPKPLFESFNDDIIF SVIYENNDQLEDVLLHVEALAKOYERASLILPNNSPREHYORQPSKEATYAH IGALFINFNRLSDSYONLSQITIRNDPTLLEIVENLRKLRNLTLOOTIINM SKATYIILSKLRNQAQIHYHNSTKDMEKTLSPQLEKVEFKNDQEEALNLEIPN DSPDLLILKTLNIFNKSILKTNFETTRKVALSFRLDSLVTKFEYETPGIFVVG NTFKGFHIFREDIARGIIRYCSRNODIYDLSKNVLDENYOLASTOORKKDDIPEG SKGYILLNPGIYVEDHDPFARFQVADAMIDILINDPLKENYVNLPRELEIFPDG TAGVDATNARVARNCPWAKSFLTGKSPSLGCIPIHDEYGTSGVARYVKIYETLN LFNSTVIRKFTQGGPDGDSNELLSSPNECYLAILGSGVLCDPKQDKDELRLH ERKNIISDPTSKLSNNGEYVSVDAMDILPNGIYVANGTEFRNTHQIIFKFDVHI FVPCGGRPNSTIILNNLHYFDEKTKCKIPIYVGANLFIPOAKNMLEBGCILFED ASANKGVSTSSMEVLASLALNDNDPFAKFIQVSGERSALYKGYVVOESRIQKME LEFGQIMNLNOLNTHISETISNOLSFITNKNLNDLVASOEMLNDLRLNVLDDKII PKIILVAGPOSVLENTPESILKVLSSYLSSTVYONGIDVNIKLELFIIGLKREA EASA"
complement(7099..9198)
/note="ORF D1014"
/codon_start=1
/protein_id="CA67476.1"
/db_xref="PID:e253334"
/db_xref="PID:91429342"
/db_xref="GI:1429342"
/db_xref="SPTREMBL:Q12310"
/translation="MSLSRLRYNRNNKTTASTLAERAYSDNAYSVSLDPTSGV MMAKTGEALNKSYDVSFSLPVADSVPRDFTSSSDDEDTDOKLTSSMEKIDIK MPENIKSIDSNIISSPAYSRKEAFIYPKGRKLRNSDPTSEFSSNLSRSTERYLN FTSKNSSTSRKLNLANDISDMKHNNSNTIGHIIPVTAATDESLACSTIEKLVN YEEERMFTRYVSLSDSVSLSTPLDQTYEAVQVTRHILEDSTAHIRKHSYTS LSSIKRLFKITFSNNNSNSCDSHPEYADCAVSSISKETSPVSTGSPSLIENE DSDRDOIOLALYSNIEASTDLVSRKYDLVYLEGSGGKYLQVRLDNKVFALKEV BSKRRRESERKVIKNTISEVCIASLKNPNICETLEIIEYKGIPOILEYCEYFLSL VMSERNAHEICIFKQILNCKYTLHDIGLSHRDLKDKCYTRRGITKLIDPRASVY FHYPLSSOMELANGIVGSDPYLSPVEVYFENYDRALDVMVGIIIFCMITRRPMK PKVDDVOFKACSGSGVSEFDLVTRATDSDSNNDYDNDGVEGVIIIDGPMFIIARLPE ETRKIMRILIEVSFFRITINGIIDDQWIKIEIFCOVVAASPNEASLRILINKNHHI TINORVAHIGLHOR"
complement(9942..10619)
/note="ORF D1018"
/codon_start=1
/protein_id="CA67478.1"
/db_xref="PID:e275392"
/db_xref="PID:91628361"
/db_xref="GI:1628361"
/db_xref="SPTREMBL:Q07623"
/translation="MGSSEDKLTKKOLAKOQFRKSEKQDOEKVKEQAPCKRPN SAAQNDGEPPVKKRKRTRGRGKNGKNGRNIIVVGSIPROITAVELONHKNS PDQRLRADKCIAPLEPADKDRPGIORMDIALLGHTLKEKINVELVGGGNS OERLEKLNKNIKIDEBERKELTLMINDGNQKIAKTATATAOTSGTDNKKVPVAGIHP DRAFLK"
10237..10539
/note="ORF D1019"
/codon_start=1
/protein_id="CA67477.1"
/db_xref="PID:e253335"
/db_xref="PID:91429343"
/db_xref="GI:1429343"
/db_xref="SPTREMBL:Q00590"

Query Match	Best Local Similarity	51.4%	Score 21.6;	DB 7;	Length 39090;
Matches	19;	Conservative	11;	Mismatches	7;
Indels	0;	Gaps	0;	Indels	0;
6	YGAYTRGNCARGAARATGCCAATTCADATRTGRA	42			
Db	38534	TGAATACGTGAGAAAAAATTCGAAATCAAGTGAA	38498		
RESULT 15					
SCDNAIV/C					
LOCUS					
SCDNAIV					
35383 bp					
DNA					
PLN					
15-APR-1996					

DEFINITION	Saccharomyces DNA for ORFs from chromosome IV.
ACCESSION	X83276
VERSION	91004294
KEYWORDS	X83276.1 GI:1004294 6-O-methylguanine-DNA methyltransferase; ADP-ribosylation factor 2; anti-silencing protein; ARF2 gene; ASF2 gene; Arpase inhibitor; glucose transport protein; high-affinity glucose transport protein; INH1 gene; methyltransferase; MGN1 gene; SHM1 gene; SNF3 gene; suppressor.
SOURCE	baker's yeast.
ORGANISM	Saccharomyces cerevisiae
REFERENCE	Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
AUTHORS	1 (bases 1 to 35383)
TITLE	Verhaaselt,P., Voelt.M. and Volckaert,G. New open reading frames, one of which is similar to the nifv gene of Azotobacter vinelandii, found on a 12.5 kbp fragment of chromosome IV of Saccharomyces cerevisiae
JOURNAL	Yeast 11 (10), 961-966 (1995)
MEDLINE	96021607
REFERENCE	2 (bases 1 to 35383)
AUTHORS	Volckaert,G.
TITLE	Direct Submission
JOURNAL	Submitted (07-DEC-1994) G. Volckaert, Lab of Gene Technology, Belgian University Leuven, Willem De Croylaan 42, 3001 Leuven.
COMMENT	Sequence overlapping with those under the following accession numbers: X58857, X56267, M60317, J054409, M21609, X16065, X23168, D004443, J03276, M82913, 023328, X60368, M94227, U08352, L07649, J03246.
FEATURES	
Source	Location/Qualifiers 1..35383 /organism="Saccharomyces cerevisiae" /strain="Fy1679" /db_xref="taxon:4932" /chromosome="IV" /clone_1tb="B. Dujon" /clone-cosmid 3184" /map="left arm" complement(1..404) /gene="D1204" /db_xref="SGD:S0002359" complement(<1..404) /gene="D1204" /codon_start=1 /db_xref="SGD:S0002359" /product="06-methylguanine DNA repair methyltransferase" /protein_id="CAA58247.1" /db_xref="PID:g1004295" /db_xref="GI:1004295" /db_xref="SWISS-PROT:P26188" /translation="MKELLYTFIEBVGALYFVEKTONLVFASIGNDKLLGKLV EGFLKREKODTMDVLDLKLAEATYKRSINVTICLENKMPDPSGALPFFELGTDFQ RKVWNLNLEHGVVYDGIARICKPTAAS" complement(627..2630) /gene="D1209" complement(627..2630) /gene="D1209" /note="putative ORF" /codon_start=1 /protein_id="CAA58248.1" /db_xref="PID:g1004296" /db_xref="GI:1004296" /db_xref="SPTRMBL:012407" /translation="MKPPLMSRSNKPLOTEANSAHADRAHOLAODFNSKODDTAIF SLPKNADIFRFENNITAHSSRSGSLYRDSADAVVLPISHTPRLSDDPYROLLODA EISLGRKKRRHSRLRTAFISFVLYVSSLSGIDGLSGNWTLSFOKFFYPLTS PLGIVATVINGAFASLAEVYSGILVEPCSRKKLOISTMTYISGAIVOTALNOMCL LGRRLVGMGPAFMSWYIIOPEPLPCIRKRLTISICVSVIASFQIWMICARY LGRPMRPSTHYALVGIITLMSFYLLESPSMLKOKNVEVALYLSNVEFDGNPEE NOTOLKRYALRKDLILKSLHQKNSYPAIYLKQFSSITKLILGOLTRINGADAFY YSPILLOMGREGKRSYITGLNALVYISVLAVYPLVLRKREKTVNLGSYVMA LPIISFDWPKSTRTYRISILAEVFLTEHFIISNDISIGWWTIELPLISAPVILVLS
gene	
CDS	
gene	
CDS	

gene
complement(2986. .3888)
/gene="D1214"
complement(2986. .3888)
/gene="D1214"
/codon_start=1
/product="suppressor of abf2"
/protein_id="CA58249.1"
/db_xref="pid:g1004297"
/db_xref="gi:1004297"
/db_xref="SWISS-PROT:p38988"
/translation="MPHDKQSGGLARLIGSASAGIMELAVFHPVDITSKRMNSHNRK
ITSGEIMNVIYFRDSEPLGRFLTPLEGYLAASYVLQRYKVGQGPANFLNRK
HKKQELNFGERTGKAMRSAAGSLIGELVYLIDLVLTIKOTNDESGKGFPIK
ILRDGSEIYVRGMGTWRTARPNRPFALPGNAPAKETIIGLKYQSQTWQONTISSI
VACGSLIYSAFLVDYIKTRIQNNRPNDESGRLVTKNLKNGVAFKGLPLPLTTT
GKLVFSPLAASLI:PRFDNLKS"
/gene="D1219"
complement(4254. .5831)
/db_xref="SGD:S0002356"
complement(4254. .5831)
/gene="D1219"
/codon_start=1
/db_xref="SGD:S0002356"
/product="anti-silencing protein"
/protein_id="CA58250.1"
/db_xref="pid:g1004298"
/db_xref="gi:1004298"
/db_xref="SWISS-PROT:p32448"
/translation="MPKRGVLDATYRSVIDSGDQSSSEFNSDREXAAVYKGLSSSR
YOKSLSRMKSEKVELOSIVHYRENAALVSSAKLIGETIGETIMAGLHGKMS
ILDENNAIKETHESSAEKRIEYLRPSKREKRNDEYLLVNLKEIKCAIKADQKNQY
MYNVAKDIFTHFEKRYEGLESLDSEVFDSESKELAKYRELNRRKSEIKVNNK
LMQVLQAKSLHLEKQAKADYDLKCIPELVDAKNTLSTKSLANORETLEAQL
ALISOOSETEKGIETETQNOVQIPNVVLVDFEENPEDLFAIOEIDLRHKKML
ADERTKNTLHERKNTIKLQSYQSLIQRLPAPQKHHGLGIFOKGSGKSCPLAPK
AAVSTYAPLLTLOSNSHOEINTPQRLLIAPADQSYSEKSTLNLDSRKYSLSR
LQPIYANIKLSTLTIKTLKTIPRPVDSQQLSRSKQSQETAKKDRPLVANEPTLDTSTIPP
PVAQSLADSKHSGSLHK"
/gene="D1224"
6078. .6407
/gene="D1224"
6078. .6407
/gene="D1224"
/note="putative ORF"
/codon_start=1
/protein_id="CA58251.1"
/db_xref="pid:g1004299"
/db_xref="gi:1004299"
/db_xref="SPTREMBL:Q12187"
/translation="MPSEKRLSKKCAATYRLVENVYICFTFYVDFCLIPRSSID
MNSGSCYFIIDFFHLSPIPWFIYFVKNHHTNISIRPTKSEKNQVSDPLFY
SCSRVAF"
/gene="D1229"
6545. .10366
/db_xref="SGD:S0002354"
6545. .10366
/gene="D1229"
/note="putative ORF"
/codon_start=1
/db_xref="SGD:S0002354"
/protein_id="CA58252.1"
/db_xref="pid:g1004300"
/db_xref="gi:1004300"
/db_xref="SWISS-PROT:p38968"
/translation="MYKLAFFSRATAPFAMSHDKLPILVSGTVSGVADNPSTDSLELE
WSLALAMQSLIAYHFAASGSNFAISNDLAKRKYTHLSTSPNSGIKQQLSVENH
RESNHSKRYTKFNNAKQDNYLASCNGNGETIFRIMNCKTSPSPYITLPGQSMSSY
DEVTSILAMQSLIAYHFAASGSNFAISNDLAKRKYTHLSTSPNSGIKQQLSVENH
PANSFRVATATSDNDPSILIMDLRNANTPLQTLNQHQKILSLDMCHQDHELFLSS
GNDYINLWPNPSAQLSQFPAKGNMCKRTFAPAPDLFACASFDNKIEVQLONTLPL

NTLDEQETFEIKQOSEETIDFRNNVNSRESKKEPVSFHLQAPTWIGESPAAHNAFGKTL
VQITPDGNGVSIITNPKISIGLESNTNTLSEALKTDKFKPLINQRLKTVLDIVNDEEDMLT
EKLSDGGEFEELKEALAFDNDESDOADDANNERKDEDEEFOQLETFNFCQEGPQSLWG
NLEQIISLNUNGVNKSISAVKNSLNDLMLMEAMVALDSSNNERLKESTKNAVPKNGSK
SLSLITVYSISKREYDQVLENLDVQOKKFTSKATONLYPNDIDQORNMMLIKIDDRKE
NGHRDQSLITLILAGSLDKVASITLSEPPLEDKLDKDKNTITEAHSECLITEERT
VFSNINSSSTINNEQILAKLEFLINLTITSGNELATEFLNSLPSNNEVYKTEKAV
ILASQSKLPAQNPATITAKSKAKTYNKTINKTSVNLPLPPGAPSTISITISMAQPGMTP
GASNAALPEKPEVPATITISAPVHEGKAPPSQOSMASPVRKNSSTIRLNSAPGMP
PYATVATVATVSTISIPONTFAPIQDGMPLMDYNQSSIPROPIINVSOTPL
NRKADGNMDLPLKYKEKPSAKAVSVAPRILISTPPLPLNGIPANMASTPAPPLSRA
PSSVMSNPBPPLPHKNSRVPSLVATSESPRASISVPAPPOSSQOFPGLISTANQJTN
TAQVSSNMPVAPPPQOBRVATPLSGVGPAPAPLRKSNISVPATAPLTOPMNGSPYPTGPT
NHTHTISPPVFNKPKPTPISTISMKRKSNTKALISQONSOGATPILSSAPGLAPPS
OPLTASIOSVNSAENVSHETPADQOPLDYLKEALVPTPLTPREYSKOLKDDKRLK
ILFYLHEQODLITOPTIDCLDLVALMKKEKKYKRLAMVITHANITATNHAQEGNMWLTGK
NLIGIAEATLN
11115. 13571
/gene="D1234"
/db_xref="SGD:S0002353"
11115. 13571
/gene="D1234"
/db_xref="SGD:S0002353"
11115. 13571
/gene="D1234"
/codon_start=1
/db_xref="SGD:S0002353"
/product="High affinity glucose transport protein"
/protein_id="CAA58253.1"
/db_xref="GI:1004301"
/db_xref="PDB:1004301"
/db_xref="SWISS-PROT:P10870"
/translation="MTSVFTDIDISTIDONSITLSEPPPOKOSMMASICGVFVANGFL
EYDIDGLNITSMTKYAKSHVAPNHDSTFAQMSILVSELSGTFEPALAPRISDSY
GKRPITLITSTLFIETISGLSDVAGAGITLLTVGVNLSIGIGALISAVPLQXQATEL
SRGMIISTYQWALITWLVSSAVSOGVHANDASVRIPGLQVWSSFLAIGMFL
PESPRTYVLRKIDLEAKSLSEFLRGVPYHDSGLLEELVEIKATYDEASGNSFIDC
FISSSRKQTLRMFTGTALQAPFOQSGINFIFYGVNFKTGVSNYSVSPFLTVAY
NVFVPLGLFTEFEGRKTYLVNGSVLNTIANFVVALYVSGISLRTAAKYMATICFL
IYAFVPLGLFTEFEGRKTYLVNGSVLNTIANFVVALYVSGISLRTAAKYMATICFL
SIGAIFETIMGSLNMGVIVAYLVTVYETRGTLTLEIDELYIKSTGVASPRFNKDI
RALKQVQPLQRLDEDKNTFVAKRNNEDEDETPRNDPNTISGLIDHSPNKEVNSIDE
RFDITQVTEILSELPKSSGNTVAPVSLQDVPIDQTEPELIRKIVYLDNGGLNGLY
FNGNSPISDSSEEDTDEIDIGSPSGOQDSNGRLMLNYNKRPSTILANSDEEANGKE
TSDNNTAKQDLAAGKERNAPQASVITDRGGLLEPETSNIILSTLSVMAQDTNENNNI
LHSSSENAITNPVNNNDLK"
14009. 15136
/gene="D1239"
14009. 15136
/gene="D1239"
/gene="D1239"
/note="putative ORF"
/codon_start=1
/protein_id="CAA58254.1"
/db_xref="PDB:91004302"
/db_xref="GI:1004302"

```

Query Match Similarity      50.24; Score 21.6; DB 7; Length 35383;
Best Local Similarity      51.48; Pred.No.2.1e+02;
Matches      19; Conservative      11; Mismatches      7; Indels      0; Gaps      0;

QY      6 YCAAYTTCGNCARGAARAATGCCCAATTCDAATRTGAA 42
      :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5218 TGAATTAAGCTGAGAAAAAATTCGCAAAATCAAAAGTAA 5182

Search completed: September 13, 1999, 16:20:31
Job time: 4543 sec

```

Search completed: September 13, 1999, 16:20:37
Job time: 4543 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:22:05 ; Search time 133.08 Seconds

(without alignments)
80,841 Million cell updates/sec

Title: US-09-325-095-2729.SEO

Perfect score: 43

Sequence: 1 GGNAGYATTCRCGNCARGA.....RTGCCRAATCDATRTGAA 43

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	58.1	6831	1 T80200	Phosphatidyl inositol
2	21.2	49.3	4488	1 V74539	Staphylococcus aureus
3	20.8	48.4	446	1 O39702	Expressed Sequence
4	20.8	48.4	446	1 O59114	Human brain Express
5	20.8	48.4	1269	1 V12774	Metal-regulated tr
6	20.8	48.4	2027	1 V32930	Homo sapiens SDF-5
7	20.8	48.4	1960	1 V34257	Human secreted pro
8	20.8	48.4	1869	1 V34199	Human secreted pro
9	20.8	48.4	1984	1 V69384	Human ATG-1622 CDN
10	20.8	48.4	1984	1 X34662	Human secreted pro
11	20.4	47.4	2289	1 O06088	Tumour cytotoxic f
12	20.4	47.4	2187	1 O10489	Hepatic parenchyma
13	20.4	47.4	2395	1 O12397	Human hepatocyte g
14	20.4	47.4	1661	1 O12398	Human hepatocyte g
15	20.4	47.4	2187	1 O12399	Human hepatocyte g
16	20.4	47.4	2172	1 O14182	Plasminogen-like g
17	20.4	47.4	2393	1 O14038	Human hepatocyte g
18	20.4	47.4	2214	1 O15176	Human leukocyte-de
19	20.4	47.4	2199	1 O15177	Human leukocyte-de
20	20.4	47.4	2184	1 O20049	Human hepatocyte g
21	20.4	47.4	2289	1 O21066	TCF II-encoding se
22	20.4	47.4	2394	1 O22143	Human HGF gene par
23	20.4	47.4	1661	1 O22144	Human HGF gene par
24	20.4	47.4	2187	1 O22145	Complete human HGF
25	20.4	47.4	2173	1 O37308	Encodes haematopo
26	20.4	47.4	2184	1 O46040	Hepatocyte growth
27	20.4	47.4	2187	1 O45702	Vasculin endotheli
28	20.4	47.4	2187	1 O47833	Competitive inhbl
29	20.4	47.4	2187	1 O47833	Competitive inhbl
30	20.4	47.4	2184	1 O56157	Hepatocyte growth
31	20.4	47.4	2187	1 V20546	Human recombinant
32	20.4	47.4	2214	1 V33626	Human leukocyte-de
33	20.4	47.4	2199	1 V33627	Human leukocyte-de
34	20.4	47.4	2172	1 V61952	Nucleotide sequenc
35	20.4	47.4	2187	1 X15186	Nucleic acid encod
36	20.2	47.0	5340	1 O75956	S. cerevisiae scsu
37	20.2	47.0	6869	1 T47442	Saccharomyces cere
38	20.2	47.0	4434	1 X33300	Fumonisin-resistan
39	19.8	46.0	9623	1 V74477	Staphylococcus aur
40	19.8	46.0	17250	1 X13020	Enterococcus faeca
41	19.6	45.6	328	1 O26721	Encodes exon IX of
42	19.6	45.6	204	1 T25315	Human gene signatu
43	19.6	45.6	2042	1 V15144	New promoter used

44 19.6 45.6 8033 1 X13058
c 45 19.4 45.1 447 1 T27954

Enterococcus faeca
Hepatitis C virus

ALIGNMENTS

```

RESULT 1
ID T80200/c
AC T80200;
DE 22-JUN-1998 (first entry)
KW Phosphatidyl inositol 3-kinase cdk CDNA.
KW antagonist; inflammatory joint disease; cell proliferation; cancer;
KW psoriasis; restenosis; atherosclerosis; therapy; diagnosis; cpk; ss.
OS Drosophila melanogaster.
FH Key
FT CDS location/Qualifiers
PI W09731650-A1.
PD 04-SEP-1997.
PF 12-FEB-1997; U02193.
PR 29-FEB-1996; US-609049.
PA (RESC) UNIV CALIFORNIA.
PI Chen Y, Moliz L, Williams LT;
DR P-PSDB; W38757.
DR P-PSDB; W38757.
PT New isolated phosphatidyl inositol-3 kinase polypeptide - used to
PT develop products for diagnosis and therapy, particularly for
PT proliferative disorders, e.g. inflammatory joint diseases, or cancer
PS Example 1; Fig 9; 77bp; English.
CC This cDNA sequence codes for cpk (see W38757), a Drosophila
CC polypeptide that belongs to a novel class of phosphatidyl inositol
CC 3-kinases that contain a C2 domain, are capable of phosphorylating
CC a D3 hydroxyl of an inositol ring in phosphatidyl inositol and
CC phosphatidyl inositol 4-phosphate, but not in phosphatidyl inositol
CC 4,5-bisphosphate, and which are involved in cell signalling cascades
CC that control e.g. cell cycle progression and intracellular protein
CC sorting. Short fragments of cpk CDNA were obtained from a
CC Drosophila cDNA library by PCR (see T80195-96). These short
CC fragments were used to screen the cDNA library to obtain larger
CC fragments, and missing 5' ends were obtained by RACE. A
CC recombinant host cell, transfected with a vector comprising a cpk
CC nucleic acid is claimed. Novel phosphatidyl inositol 3-kinases
CC can be used to screen for agonists/antagonists of activity and
CC in a claimed method of treating a disorder caused by dysregulation
CC of a growth factor activation signalling cascade. Antagonists
CC may reduce Ras activation allowing treatment of proliferative
CC disorders such as atherosclerosis, inflammatory joint disease,
CC psoriasis, restenosis following angioplasty, and cancer.
SQ Sequence 6831 BP; 1829 A; 1685 C; 1669 G; 1648 T;

Query Match 58.1%; Score 25; DB 1; Length 6831;
Best Local Similarity 60.5%; Pred. No. 0.65;
Matches 23; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

OY 6 YGAYYTRCGNAGARATGCGRAATCDATRTGAA 43
DB 4544 TGGCATCGCCAGAACTTCCAAAGTCATATGTGAA 4507

RESULT 2
ID V74539
AC V74539;
DE 16-MAR-1999 (first entry)
KW Staphylococcus aureus contig SBO ID #228.
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scaled skin syndrome;
KW toxic shock syndrome; ds.

```

```

OS Staphylococcus aureus.
PH Key Location/Qualifiers
FT misc-feature 721..780
FT /tag-a
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc-feature 2521..2580
FT /tag-b
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc-feature 4321..4380
FT /tag-c
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 996-999; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium
SQ Sequence 4488 BP; 1698 A; 686 C; 675 G; 1238 T;

Query Match 49.3%; Score 21.2; DB 1; Length 4488;
Best Local Similarity 52.5%; Pred. No. 20;
Matches 21; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

OY 2 GNGAYGAYYTRCGNCARGARARATGCCRAARTCDATRTGR 41
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 3183 GGCATCATATTAAGCGAGACAAATTCACAAATCATCTCGA 3222

RESULT 3
ID 039702/C
AC 039702;
DT 20-MAY-1993 (first entry)
DE Expressed Sequence Tag human gene marker EST00257.
KW expressed sequence tag; human genome project; chromosome;
KW human gene sequencing; PCR mapping; somatic cell hybrids;
KW sublocalisation; gene tagging; tissue typing.
OS Synthetic.
PN MO9300353-A.

```

```

PD 07-JAN-1993.
PF 19-JUN-1992; 005222.
PR 20-JUN-1991; US-716831.
PR 12-FEB-1992; US-837195.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Venter CJ;
PI WPI: 93-036325/04.
PT Particular expressed sequence tags from human cDNA - corresponds
PT to transcription prods. of genes, useful for tagging genes.
PT mapping chromosomes and tissue typing
PS Claim 3; Page 107; 199pp; English.
CC This sequence represents an EST (expressed sequence tag) ESTs are markers
CC for human genes actually transcribed in vivo. Unlike the random genomic
CC DNA sequence tagged sites (STSs), ESTs point directly to expressed genes.
CC The use of ESTs could facilitate the tagging of most expressed human
CC genes within a few years at a fraction of the cost of complete genomic
CC sequencing. Using PCR primers 039419-039580 (sequences designed
CC from the ESTs) sublocalisation of an EST can be achieved with panels of
CC fragments from specific chromosomes or pools of large genomic clones in
CC an analogous manner. This sequence represents EST00257.
SQ Sequence 446 BP; 82 A; 116 C; 103 G; 144 T;

Query Match 48.4%; Score 20.8; DB 1; Length 446;
Best Local Similarity 47.6%; Pred. No. 14;
Matches 20; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

OY 2 GNGAYGAYYTRCGNCARGARARATGCCRAARTCDATRTGR 43
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 198 GTCACGACCTCGAGTCAAGAAAGTGCAGAAATGAGCCCGAA 157

RESULT 4
ID 059114/C
AC 059114;
DT 16-MAR-1994 (first entry)
DE Human brain Expressed Sequence Tag EST00257.
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcripction; mapping; locations; chromosomes; chromosome1; ss.
OS Homo sapiens.
PN MO9316178-A.
PF 18-AUG-1993.
PF 12-FEB-1993; 001294.
PR 12-FEB-1992; US-837195.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
PI WPI: 93-272882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
PS Example 1; Page 123; 500pp; English.
CC The expressed sequence tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.
CC EST00257 hybridises to chromosome 3 and has been putatively
CC identified as being associated with kinesin. It shows strong
CC sequence similarity to the squid (67%) and Drosophila (70.4%)
CC kinesin heavy chain. See also 059041-061440.
SQ Sequence 446 BP; 82 A; 116 C; 103 G; 144 T;

Query Match 48.4%; Score 20.8; DB 1; Length 446;
Best Local Similarity 47.6%; Pred. No. 14;
Matches 20; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

OY 2 GNGAYGAYYTRCGNCARGARARATGCCRAARTCDATRTGR 43
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 198 GTCACGACCTCGAGTCAAGAAAGTGCAGAAATGAGCCCGAA 157

```

```

RESULT 5
V12774/c
ID V12774 standard; cDNA; 1269 BP.
AC V12774;
DE 30-JUL-1998 (first entry)
DE Metal-regulated transporter gene ZMT2.
KM ss; MMT; transgenic plant; pollution; metal-deficiency.
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT CDS 1..1269
FT /tag= a
FT /product= "metal-regulated transporter protein ZMT2"
PN MO974500-A1.
PD 04-DEC-1997.
PE 27-NOV-1996; U19065.
PK 11-OCT-1996; CA-187728.
PR 29-MAY-1996; US-018578.
PA (DART-) DARTMOUTH COLLEGE.
PA (MINU) UNIV MINNESOTA.
PI Elide DJ, Gueriot ML;
DR WPI: 98-032230/03.
DR P-PSDB: W41169.
PT Novel A. thaliana metal-regulated transporter proteins and related
PT DNA - for generating transgenic plants useful for removing
PT pollutants from soil or providing nutrient to patients suffering
PT from metal-deficiency disorders
PS Disclosure: Fig 22; 144pp; English.
CC The metal-regulated transporter (MRT) genes (V12774 and V12755) encode
CC polypeptides which are capable of transporting metals such as Fe(II),
CC Pb and Zn. The peptides can be used to generate transgenic plants in
CC which the expression of MRT can be altered. These plants can be used
CC to remove pollutants from the soil or to treat a disorder associated
CC with metal-deficiency. The polypeptides can also be used to identify
CC agents which inhibit or stimulate MRT or MRT expression.
SO Sequence 1269 BP; 343 A; 259 C; 275 G; 392 T;

Query Match 48.4%; Score 20.8; DB 1; Length 1269;
Best Local Similarity 47.6%; Pred. No. 20;
Matches 20; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

OY 2 GNGAGVAYTRCGNCARARARARCCCAATCDAATTCGAA 43
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
DB 362 GTGATGATTTCAGTGAAGAAAGTAAGAAAGCGACATTAA 321

RESULT 6
V32930
ID V32930 standard; DNA; 2027 BP.
AC V32930;
DE 09-NOV-1998 (first entry)
DE Homo sapiens SDF-5 gene.
KM SDF-5; frazzled; osteoarthritis; rheumatoid arthritis;
KM cell formation; proliferation; differentiation; diabetes;
KM pancreatic cancer; wound healing; gene therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 256..1143
FT /tag= a
FT /product= SDF-5 protein
FT /tag= b
FT s1g-peptide
PN MO9835043-A1.
PD 13-AUG-1998.
PE 15-OCT-1997; U18369.
PR 08-MAY-1997; US-848439.
PR 06-FEB-1997; US-796153.
PA (GEMV) GENETICS INST INC.
PA Lavallee ER, Racine LA;
DR WPI: 98-447240/38.
DR P-PSDB: W49082.

```

```

PT Isolated DNA encoding human SDF-5 protein - useful for controlling
PT growth, differentiation etc. of cells, particularly of chondrocytes
PT for treatment of arthritis etc., also pancreatic cells
PS Claim 1; Page 55-57; 69pp; English.
CC The sequence is that encoding human SDF-5, a member of the Frazzled
CC protein family. Cells transformed with a vector containing the sequence
CC are used to regulate genes, particularly pancreatic genes, or
CC in combination with bone morphogenic protein 2 (BMP2), to
CC increase differentiation of progenitor cells into chondrocytes.
CC The protein may be used to treat osteoarthritis, rheumatoid
CC arthritis, or articular cartilage defects, also to increase/inhibit cell
CC formation, growth, differentiation, proliferation and/or maintenance in
CC many other organs or tissues, e.g. for prevention or treatment of
CC pancreatic cancer, diabetes (by inducing de novo formation of islet
CC cells), other tissue defects, also to improve healing of wounds and to
CC increase survival of nervous system cells, e.g. in cases of transplants
CC The coding sequence can be used in gene therapy, and its fragments to
CC detect related mRNA, while the protein is also used to generate
CC antibodies, useful for affinity purification and as immunoassay
CC reagents. Many other potential uses/activities for the gene and its
CC encoded are contemplated but not exemplified, e.g. as cytokines,
CC immuno-suppressants or immunostimulants, regulators of haematopoiesis,
CC as fertility-control agents, haemostatic or thrombolytic agents,
CC anti-inflammatory agents, antimicrobials, modulators of biohythms
CC and many more.
SO Sequence 2027 BP; 496 A; 583 C; 458 G; 490 T;

Query Match 48.4%; Score 20.8; DB 1; Length 2027;
Best Local Similarity 47.6%; Pred. No. 23;
Matches 20; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

OY 1 GNGAGVAYTRCGNCARARARARCCCAATCDAATTCGAA 42
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
DB 858 GGAGTATACCTACATCAACGACGAGATACCAAAATCATCTGGA 699

RESULT 7
V34257
ID V34257 standard; DNA; 1960 BP.
AC V34257;
DE 28-JAN-1999 (first entry)
DE Human secreted protein gene 46 clone HLED731.
KM Human; secreted protein; fusion protein; gene therapy; protein therapy;
KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KM developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KM immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KM inflammation; ischaemic shock; Alzheimer's disease; rescues; AIDS;
KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN MO9839446-A2.
PD 11-SEP-1998.
PE 06-MAR-1998; U04492.
PR 07-MAR-1997; US-038621.
PR 07-MAR-1997; US-040161.
PR 07-MAR-1997; US-040162.
PR 07-MAR-1997; US-040163.
PR 07-MAR-1997; US-040333.
PR 07-MAR-1997; US-040334.
PR 07-MAR-1997; US-040336.
PR 07-MAR-1997; US-040626.
PR 11-APR-1997; US-043311.
PR 11-APR-1997; US-043312.
PR 11-APR-1997; US-043313.
PR 11-APR-1997; US-043314.
PR 11-APR-1997; US-043315.
PR 11-APR-1997; US-043568.
PR 11-APR-1997; US-043569.
PR 11-APR-1997; US-043576.
PR 11-APR-1997; US-043578.
PR 11-APR-1997; US-043580.

```

[illegible]

CC murine SDF5. ATG-1622 polypeptides and polynucleotides are useful for
 CC diagnosing susceptibility to diseases by detecting mutations in the
 CC ATG-1622 gene and can diagnose diseases associated with ATG-1622
 CC imbalance. The polypeptides can be used to screen for agonists and
 CC antagonists which can be used in treatment to activate or inhibit
 CC ATG-1622 activity. The ATG-1622 polypeptide can be administered directly
 CC or as a vaccine to inoculate against disease. Diseases which can be
 CC diagnosed, prevented or treated by the ATG-1622 polypeptide or
 CC polynucleotides include heart disease, hypertension, kidney diseases,
 CC obesity, insulin resistance, lipodystrophy, diabetes and central nervous
 CC system (CNS) diseases. The ATG-1622 polypeptide is also useful for
 CC mapping the gene to a chromosome, allowing gene inheritance to be
 CC studied through linkage analysis.
 SQ Sequence 1984 BP; 477 A; 564 C; 452 G; 484 T;

Query Match 48.4%; Score 20.8; DB 1; Length 1984;
 Best Local Similarity 47.6%; Pred. No. 22;
 Matches 20; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

OY 1 GGNAGAYTTCGNCARGAARATGCCRAARTCDATRTGGA 42
 DB 841 GGAGATACCTACATCAACGAGATACCAAAATCATCTCGA 882

RESULT 10
 X34662
 ID X34662 standard; cDNA; 1984 BP.

AC X34662;
 DT 01-JUL-1999 (first entry)
 DE Human secreted protein ligand ATG-1622 polypeptide encoding cDNA.
 KW Human; ATG-1622 polypeptide; secreted protein ligand; 7-TM receptor;
 KW Fizzled family; heart disease; hypertension; kidney disease; obesity;
 KW insulin resistance; lipodystrophy; diabetes; CNS disease; ss.
 OS Homo sapiens.
 PN CA2229436-A.
 PD 21-NOV-1998.
 PF 23-APR-1998; 229436.
 PR 20-MAR-1998; US-0448865.
 PR 21-MAY-1997; US-047251.
 PR 13-JUN-1997; US-874156.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Hu E. Zhu Y;
 DR WPI: 99-244679/21.
 DR P-PSDB: Y06923.

PT New secreted protein ligand polypeptide (ATG-1622) of 7-TM (fizzled
 PT family) receptors, useful for diagnosis, prevention and treatment of
 PT obesity, lipodystrophy, hypertension and heart disease
 PS Claim 4; Page 10; 42pp; English.
 CC This cDNA encodes a human ATG-1622 polypeptide, a secreted protein ligand
 CC for 7-TM (fizzled family) receptors. ATG-1622 polypeptides and
 CC polynucleotides are useful for diagnosing a disease or susceptibility to
 CC a disease by determining the presence/absence of a mutation in the ATG-
 CC 1622 gene, and/or analysing for the presence or amount of ATG-1622
 CC expressed in the patient. ATG-1622 expressing cells are useful for
 CC identifying modulators of the polypeptide which are can be used to treat
 CC conditions associated with a lack or excess of ATG-1622 polypeptide.
 CC Diseases diagnosed, prevented or treated include heart disease,
 CC hypertension, kidney diseases, obesity, insulin resistance,
 CC lipodystrophy, diabetes and CNS diseases. ATG-1622 protein may be useful
 CC for treating or preventing the onset of obesity. ATG-1622 polypeptides
 CC are also useful for mapping genes to chromosomes, allowing gene
 CC inheritance to be studied through linkage analysis.
 SQ Sequence 1984 BP; 477 A; 564 C; 452 G; 484 T;

Query Match 48.4%; Score 20.8; DB 1; Length 1984;
 Best Local Similarity 47.6%; Pred. No. 22;
 Matches 20; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

OY 1 GGNAGAYTTCGNCARGAARATGCCRAARTCDATRTGGA 42
 DB 841 GGAGATACCTACATCAACGAGATACCAAAATCATCTCGA 882

RESULT 11

ID 006088 standard; cDNA; 2289 BP.
 AC 006088;
 DT 31-JAN-1991 (first entry)
 DE Tumour cytotoxic factor II gene.
 KW TCF II; glycoprotein; fibroblasts; anticancer agent; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 78..2249
 FT /tag- a

PN W09010651-A.
 PD 20-SEP-1990.
 PF 03-SEP-1990; J00314.
 PR 03-OCT-1989; JP-058631.
 PR 16-JAN-1990; JP-006692.
 PA (SNOW) SNOW BRAND MITK PRODUCTS.
 PI Higashio K; Mistuda S; Shima N; Itagaki Y; Nagao M.
 DR WPI: 90-304990/40.
 DR P-PSDB: R07144.

PT Human glycoprotein having antitumour activity - for inducing
 PT differentiation of leukemia cells and enhancing cell mediated
 PT immunity.
 PS Disclosure: Fig 15; 73pp; Japanese.
 CC The sequence was obtd. from a clone isolated from a cDNA library
 CC prep'd. from RNA ext'd. from fibroblast IMR-90 cells (ATCC CCL-186).
 CC The DNA can be used to produce the TCF II by recombinant DNA
 CC technology. TCF II is an anticancer agent and induces
 CC differentiation of leukemia cells, enhances cell-mediated immunity,
 CC and accelerates the proliferation of human blood vessel endothelial
 CC cells and hepatic parenchymal cells.
 SQ Sequence 2289 BP; 703 A; 491 C; 523 G; 572 T;

Query Match 47.4%; Score 20.4; DB 1; Length 2289;
 Best Local Similarity 52.5%; Pred. No. 34;
 Matches 21; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

OY 2 GGNAGAYTTCGNCARGAARATGCCRAARTCDATRTGGA 41
 DB 1099 GCAAGACCTACGAGAAATTTACTGCCGAATCCAGATCG 1138

RESULT 12
 ID 010489 standard; DNA; 2187 BP.

AC 010489;
 DT 16-APR-1991 (first entry)
 DE Hepatic parenchymal cell growth factor gene.
 KW Hepatic parenchymal cell growth factor; cirrhosis; ss.

OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 1..2187
 FT /tag- a
 FT /product- hHGF
 FT misc_rna 88..2187
 FT /tag- b
 FT /product- hHGF active fragment
 FT 94..2187
 FT /tag- c
 FT /product- hHGF active fragment

PN EP-412557-A.
 PD 13-FEB-1991.
 PF 10-AUG-1990; 115397.
 PR 11-AUG-1988; JP-209449.
 PR 03-APR-1990; JP-088392.
 PR 27-JUL-1990; JP-200898.
 PA (MIRU) MITSUBISHI KASEI CORP.
 PI Kitamura N; Miyazawa K; Daikuhara Y; Tsudouchi H; Naka D;
 PI Takahashi, Matsui R, Yoshiyama Y;

DR WPI: 91-045716/07.
 DR P-PSDB: R10656.
 PR Hepatic parenchymal cell growth factor - produced by recombinant
 PT methods, useful for e.g. treating cirrhosis
 PS Disclosure; fig 2; 46pp; English.
 CC This hepatic growth factor gene and its active fragments (see
 CC feature table) are contained within the expression plasmid,
 CC PKCHGF-2. This is used to transform host cells for recombinant
 CC hHGF is useful in the treatment of hepatic diseases, e.g. cirr-
 CC hosis. It is produced in large amounts, stably using this method.
 SQ Sequence 2187 BP; 673 A; 456 C; 501 G; 557 T;

Query Match 47.4%; Score 20.4; DB 1; Length 2187;
 Best Local Similarity 52.5%; Pred. No. 33;
 Matches 21; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

OY 2 GNGATGAYTTCGNCARGAARATGCCRAATCDAATRTGR 41
 DB 1037 GCAAGGACCTACGAGAAATTAATCTGCGGAATCCAGATGG 1076

RESULT 13
 ID Q12397 standard; cDNA; 2395 BP.
 AC Q12397;
 DT 17-SEP-1991 (first entry)
 DE Human hepatocyte growth factor HBC25 cDNA.
 KM Human hepatocyte growth factor; HGF; recombinant; HBC25; ss.
 PN J03130091-A.
 PD 03-JUN-1991.
 PF 05-JUN-1989; 142697.
 PR 05-JUN-1989; JP-142697.
 RA (TOYM) TOYORO KK.
 DR WPI: 91-204688/28.
 PT Recombinant human hepatocyte growth factor - used to produce
 PT polypeptide which promotes in-vitro growth of hepatocyte(s)
 PS Disclosure; fig 3(a); 20pp; Japanese.
 CC The complete sequence for the recombinant hHGF is presented in
 CC Q12399. The GF can be used in compns. for the treatment of
 CC hepatic disorders. See also Q12397-99.
 SQ Sequence 2395 BP; 748 A; 458 C; 501 G; 688 T;

Query Match 47.4%; Score 20.4; DB 1; Length 2395;
 Best Local Similarity 52.5%; Pred. No. 34;
 Matches 21; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

OY 2 GNGATGAYTTCGNCARGAARATGCCRAATCDAATRTGR 41
 DB 465 GCAAGGACCTACGAGAAATTAATCTGCGGAATCCAGATGG 504

RESULT 14
 ID Q12398 standard; cDNA; 1661 BP.
 AC Q12398;
 DT 17-SEP-1991 (first entry)
 DE Human hepatocyte growth factor HAC19 cDNA.
 KM Human hepatocyte growth factor; HGF; recombinant; HAC19; ss.
 PN J03130091-A.
 PD 03-JUN-1991.
 PF 05-JUN-1989; 142697.
 PR 05-JUN-1989; JP-142697.
 RA (TOYM) TOYORO KK.
 DR WPI: 91-204688/28.
 PT Recombinant human hepatocyte growth factor - used to produce
 PT polypeptide which promotes in-vitro growth of hepatocyte(s)
 PS Disclosure; fig 3(b); 20pp; Japanese.
 CC The complete sequence for the recombinant hHGF is presented in
 CC Q12399. The GF can be used in compns. for the treatment of
 CC hepatic disorders. See also Q12397-99.

SQ Sequence 1661 BP; 526 A; 377 C; 363 G; 395 T;

Query Match 47.4%; Score 20.4; DB 1; Length 1661;
 Best Local Similarity 52.5%; Pred. No. 31;
 Matches 21; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

OY 2 GNGATGAYTTCGNCARGAARATGCCRAATCDAATRTGR 41
 DB 1064 GCAAGGACCTACGAGAAATTAATCTGCGGAATCCAGATGG 1103

RESULT 15
 ID Q12399 standard; DNA; 2187 BP.
 AC Q12399;
 DT 17-SEP-1991 (first entry)
 DE Human hepatocyte growth factor.
 KM Human hepatocyte growth factor; HGF; recombinant; ss.
 PN J03130091-A.
 PD 03-JUN-1991.
 PF 05-JUN-1989; 142697.
 PR 05-JUN-1989; JP-142697.
 RA (TOYM) TOYORO KK.
 DR WPI: 91-204688/28.
 DR P-PSDB: R12792.
 PT Recombinant human hepatocyte growth factor - used to produce
 PT polypeptide which promotes in-vitro growth of hepatocyte(s)
 PS Disclosure; fig 4(1-4); 20pp; Japanese.
 CC The GF can be used in compns. for the treatment of
 CC hepatic disorders. See also Q12397-99.
 SQ Sequence 2187 BP; 676 A; 460 C; 500 G; 551 T;

Query Match 47.4%; Score 20.4; DB 1; Length 2187;
 Best Local Similarity 52.5%; Pred. No. 33;
 Matches 21; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

OY 2 GNGATGAYTTCGNCARGAARATGCCRAATCDAATRTGR 41
 DB 1037 GCAAGGACCTACGAGAAATTAATCTGCGGAATCCAGATGG 1076

Search completed: September 13, 1999, 16:22:08
 Job time: 4028 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:14:01 ; Search time 849.94 Seconds

(without alignments)
99.794 Million cell updates/sec

Title: US-09-325-095-2729.SEQ

Perfect score: 43
Sequence: 1 GGGAGATGATTCGNCARCA.....RTGCCAATTCATRTGAA 43

Scoring table:

Search: 2546578 seqs, 98626752 residues

Database :

EST: *
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	22.8	53.0	289	44	AA728172	AA728172 Lmlv39p7/
2	22	51.2	384	27	AA028985	AA028985 ZK11C01.r
3	22	51.2	599	38	AA801073	AA801073 EST190570
4	22	51.2	572	51	AI726663	AI726663 BRLG1578
5	21.8	50.7	458	23	H51655	H51655 YP81B08.r1
6	21.8	50.7	462	26	W77968	W77968 zd70e11.r1
7	21.8	50.7	641	48	AI559028	AI559028 AEMTBA58
8	21.6	50.2	474	33	AA391489	AA391489 LD10161.5
9	21.4	49.8	392	34	AA499435	AA499435 v185e01.r
10	21.4	49.8	378	34	AA512088	AA512088 v143n01.r
11	21.4	49.8	245	37	AA725990	AA725990 vU84h11.r
12	21.4	49.8	579	37	AA726611	AA726611 vU93d09.r
13	21.4	49.8	638	37	AA726685	AA726685 vU93h06.r
14	21.4	49.8	600	38	AA795990	AA795990 vs66d10.r
15	21.4	49.8	450	39	C88410	C88410 C88410 Carp
16	21.4	49.8	719	40	AA851519	AA851519 LD32028.5
17	21.4	49.8	551	40	AA951564	AA951564 LD32147.5
18	21.2	49.3	490	20	T41916	T41916 5179 Lambda
19	21.2	49.3	435	27	W99426	W99426 TgESTz201d1
20	21.2	49.3	460	35	C28257	C28257 C28257 Rice
21	21.2	49.3	460	35	C28712	C28712 C28712 Rice
22	21.2	49.3	433	39	AA856072	AA856072 w82e08.r
23	21.2	49.3	340	46	AI433117	AI433117 tN41f05.x
24	21.2	48.8	444	51	AI737481	AI737481 T110108e
25	20.8	48.4	446	20	M62196	M62196 EST00257.H1
26	20.8	48.4	406	23	H16121	H16121 Y128C05.r1
27	20.8	48.4	359	31	AA295018	AA295018 EST100474
28	20.8	48.4	302	32	AA330758	AA330758 EST34493
29	20.8	48.4	378	33	AA429960	AA429960 zW8810.r
30	20.8	48.4	444	37	AA706119	AA706119 z159503.s
31	20.8	48.4	515	42	AI105743	AI105743 SWAMCACC25
32	20.6	47.9	416	20	D24386	D24386 RICR1821A.R
33	20.6	47.9	429	20	D24456	D24456 RICR1950A.R
34	20.6	47.9	330	20	Z33790	Z33790 ATR52881.Ye
35	20.6	47.9	430	23	T75952	T75952 10730 Lambda
36	20.6	47.9	544	24	N38187	N38187 19414 Lambda
37	20.6	47.9	334	35	C28727	C28727 C28727 Rice
38	20.6	47.9	462	36	C72767	C72767 C72767 Rice
39	20.6	47.9	287	40	C91804	C91804 C91804 Rice
40	20.6	47.9	327	50	AU064480	AU064480 AU064480
41	20.6	47.9	353	50	AU064483	AU064483 AU064483
42	20.6	47.9	260	51	AU068604	AU068604 AU068604
43	20.4	47.4	491	21	R13658	R13658 Yf60f07.r1
44	20.4	47.4	448	22	R52797	R52797 Yg99f12.r1
45	20.4	47.4	522	29	AA156676	AA156676 z118h04.S

ALIGNMENTS

RESULT 1
AA728172 289 bp mRNA EST 10-DEC-1998
LOCUS Lmlv39p7/352M Leishmania major RNA
DEFINITION library from logarithmic stage (day 7) Leishmania major cDNA clone
352M 5', mRNA sequence.
ACCESSION AA728172
NTD 93926020

VERSION AA728172.1 GI:3926020
 KEYWORDS EST
 SOURCE Leishmania major
 ORGANISM Leishmania major
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania
 1 (bases 1 to 289)
 Almeida, R., Sampato, I., Schneider, H. and Blackwell, J.M.
 Analysis of Leishmania major promastigote library from different
 stages of development
 Unpublished (1998)
 JOURNAL
 COMMENT On Jan 5, 1998 this sequence version replaced gi:2747129.

CONTACT: Blackwell JM
 Cambridge Institute for Medical Research
 Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road,
 Cambridge CB2 2XY, UK
 Tel: 01223 336 143
 Fax: 01223 331 206
 Email: jmb37@cus.cam.ac.uk
 PCR Primers
 FORWARD: GTAAACGACGGCCAGT
 BACKWARD: GGAACAGCATGACATG
 Seq primer: AATTACCTCCTACTAAGG
 High quality sequence stop: 289.
 Location/Qualifiers
 1..289
 /organism="Leishmania major"
 /strain="LV39"
 /db_xref="taxon:5664"
 /clone="352W"
 /clone_1lb="Leishmania major promastigote full length cDNA
 library from logarithmic stage (day 7)"
 /cell-type="Promastigote"
 /note="Vector: Lambda Zap II; Site_1: XhoI; Site_2: NotI"

BASE COUNT 62 a 85 c 85 g 57 t
 ORIGIN

Query Match 53.0%; Score 22.8; DB 44; Length 289;
 Best Local Similarity 52.4%; Pred. No. 46;
 Matches 22; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

OY 1 GGNGATGAYTTCGNCARBARAARTGCCAARTCCATRTGGA 42
 DB 55 GGTGACAGCTTCGCGCAGAGAAATCGGAAATCGCGTGA 96

RESULT 2
 LOCUS AA028985 384 bp mRNA EST 19-NOV-1996
 DEFINITION zkl1c01.1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
 IMAGE:470208 5' similar to gb:x74929 KERATIN, TYPE II CYTOSKELETAL
 8 (HUMAN)), mRNA sequence.
 ACCESSION AA028985
 NID 91496387
 VERSION AA028985.1 GI:1496387
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 384)
 Hillier, L., Clark, N., Dubugue, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevas, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)

TITLE
 JOURNAL
 COMMENT
 Contact: Wilson RK
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LINT; contact the
 IMAGE Consortium (info@image.lnln.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 349.
 Location/Qualifiers
 1..384
 /organism="Homo sapiens"
 /db_xref="GDB:3756210"
 /db_xref="taxon:9606"
 /clone="IMAGE:470208"
 /clone_1lb="Soares_pregnant_uterus_NbHPU"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
 Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer [5',
 AACTGGAAGATTCGCGCGCCGCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by M. Fatima Bonaldo."

BASE COUNT 105 a 94 c 119 g 66 t
 ORIGIN

Query Match 51.2%; Score 22; DB 27; Length 384;
 Best Local Similarity 47.6%; Pred. No. 97;
 Matches 20; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

OY 1 GGNGATGAYTTCGNCARBARAARTGCCAARTCCATRTGGA 42
 DB 289 GGGATGACCTCGCGCGACAAAGACTGAGATCTGTGAA 330

RESULT 3
 LOCUS AA801073 599 bp mRNA EST 30-APR-1998
 DEFINITION EST190570 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
 ROVA0A34 5' end, mRNA sequence.
 ACCESSION AA801073
 NID 92864028
 VERSION AA801073.1 GI:2864028
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 599)
 Lee, N.H., Gloder, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 JOURNAL
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1398005.

TITLE
 JOURNAL
 COMMENT
 Contact: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..599
 /organism="Rattus sp."
 /db_xref="ATCC (lnhost):2008443"

FEATURES
 source

```

/db_xref="taxon:10118"
/map="12g"
/clone="ROVAA34"
/clone_id="Normalized rat ovary, Bento Soares"
/node="Organ: ovary; Vector: pT73Pac; Site_1: ECORI,
Site_2: NciI"
BASE COUNT      154 a      162 c      157 g      126 t
ORIGIN

```

Query Match	51.2%	Score 22	DB 38	Length 599
Best Local Similarity	47.6%	Pred. No.	1e+02	
Matches	20	Conservative	12	Mismatches 10; Indels 0; Gaps 0;

```

Oy      1  GGNGAYGAYTTRCGNCARBARAARTGCCRAARTCDTARTGGA 42
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      359  GGAGATGACTTACTCAGAGAGAACCAAGAGGGAATGAA 400

```

RESULT	4
LOCUS	A1726663/c
DEFINITION	A1726663 572 bp mRNA EST 11-JUN-1999 BNHG151782 Six-day Cotton fiber <i>Gossypium hirsutum</i> cDNA 5' sml11811

LOCUS	572 bp	MRNA	EST	11-JUN-1999
DEFINITION	AI726663			
	BNHG15782	Slr-day Cotton fiber Gossypium hirsutum cDNA 5' sln11rnt		
		TO ADENYLOSUCCINATE SYNTHETASE PRECURSOR (IMP-ASPARTATE LIGASE)		
	g111616657 (U49389)	adenylosuccinate synthetase [Arabidopsis thaliana], mRNA sequence.		
ACCESSION	AI726663			
	AI726663			

ACCESSION	AI726663
NID	95045515
VERSION	AI726663.1
RECORD	GI:5045515

ORGANISM

REFERENCE
1 (bases 1 to 572)

REFERENCE	1 (bases 1 to 572)
AUTHORS	Brewitt, M., Metz, E.C., Davy, D.F. and Burr, B.
TITLE	ESTs from developing cotton fiber
JOURNAL	Unpublished (1999)
COMMENT	On May 18, 1998 this sequence version replaced g1:3138270.

FEATURES

Contact: Ben Burr
Biology Department
Brookhaven National Laboratory
Upton, NY 11973, USA
Tel: 516-344-3396
Fax: 516-344-3407
Email: burrsbnl@bnl.gov
Seq primer: 53 Primer.
Location/Qualifiers

```

SOURCE
1. 5/2
/organism="Gossypium hirsutum"
/cultivar="Acala Maxxa"
/db.xref="taxon:3635"
/map="18"
/clone_lib="Six-day Cotton fiber"
/tissue_type="immature fiber"
/dev_stage="Six days post anthesis"
/lab_host="X11-blue"
/note="Vector: pInescript II KS+"
128 a 117 c 142 g 162 t 3 others
BASE COUNT
ORIGIN

```

Query Match	51.28;	Score 22;	DB 51;	Length 572;
Best Local Similarity	57.18;	Pred. No. 1e+02;		
Matches 20;	Conservative 9;	Mismatches 6;	Indels 0;	Gaps 0;

```
QY      8 AYYTRGCNCARGAARATGCCRAARTCDATRGRGA 42
          | : : | : : | : : | : : | : : | : : |
Db     64 AGTTACGAAAGCATAGTACCAGAATCATGTTCGA 30
```

RESULT	5
H51655/c	
LOCUS	H51655
DEFINITION	Y818108.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:193815 5' similar to contains Alu repetitive element,, mRNA sequence.
ACCESION	U00001

ACCESSION	H51655
NID	g991496
VERSION	H51655.1
KEYWORDS	GI:991496
SOURCE	EST.
ORGANISM	human.
	homo sapiens

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia	
Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
1 (bases 1 to 458)	
REFERENCE	
AUTHORS	Hallier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M.,

TITLE	The WashU-Merck EST Project
JOURNAL	unpublished (1995)
COMMENT	On Sep 21, 1992 this sequence version replaced g1:276010

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estelw@wustl.edu

High quality sequence strops: 334
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 708 Std Error: 0.00
Seq primer: M13RPI
High quality sequence strop: 334.

FEATURES

SOURCE

```

/organism="Homo sapiens"
/db_xref="GDB:3762836"
/db_xref="taxon:9606"
/clone="IMAGE:193815"
/clone_id="Soares fetal liver spleen INF15"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notice="Organ: Liver and Spleen; Vector: pT7n3D (Pharmacia) with a modified polylinker. Site.1: Pac I; Site.2: Eco RI. 1st strand cDNA was primed with a Pac I - oligo(dT) primer 15', ACCTGGAGAGATTAATTAAAGACTCTTTTCTTTTCTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7n3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonalodo."

```

Query Match:	50.7%	Score 21, 8:	DB 23,	Length 458;
Best Local Similarity	55.3%	Pred. No. 1.2e+02;		
Matches 21; Conservative	9;	Mismatches 8;	Indels 0;	Gaps 0;

Qy 5 AYGAYYTRCGCARGAABAARTGCCRAARTCDATRIGRA 42
| | | : : | | | : | | | : | | | : |
Db 145 AAGATTGCAGCATGAATGCGCCCAATATATATANGA 108

RESULT	6			
W77968				
LOCUS	W77968	462 bp	mRNA	EST
				17-OCT-1996

/lab_host="SOLR (kanamycin resistant)"
 /note="Organ: Skin; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
 sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

Query Match 49.8%; Score 21.4; DB 34; Length 378;
 Best Local Similarity 46.5%; Pred. No. 1.5e+02;
 Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

Db 1 GGNAGAYTTCRCNARGARARATCCRAATCDATRTGAA 43
 192 GGAGATGACTTCTCAAGAGAACCCAGAGAGAAATGAAA 234

RESULT 11
 LOCUS AA725990 245 bp mRNA EST 02-JAN-1998
 DEFINITION vus4h11.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
 IMAGE:1198149.5, similar to gb:J00124.cd61 KERATIN, TYPE I
 CYTOSKELETAL 14 (HUMAN); gb:M13806 Mouse Keratin (MOUSE);, mRNA
 sequence.

ACCESSION AA725990
 MID 92743697
 VERSION AA725990.1 GI:2743697
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 245)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)

TITLE
 JOURNAL
 COMMENT On May 18, 1995 this sequence version replaced gi:811118.

CONTACT: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:645245
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 173.

FEATURES

source
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:1198149"
 /clone_1lb="Stratagene mouse skin (#937313)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor

BASE COUNT 69 a 66 c 74 g 36 t
 ORIGIN

Query Match 49.8%; Score 21.4; DB 37; Length 245;
 Best Local Similarity 46.5%; Pred. No. 1.5e+02;
 Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

Db 1 GGNAGAYTTCRCNARGARARATCCRAATCDATRTGAA 43
 60 GGAGATGACTTCTCAAGAGAACCCAGAGAGAAATGAAA 102

RESULT 12
 LOCUS AA726611 579 bp mRNA EST 02-JAN-1998
 DEFINITION vus93d09.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
 IMAGE:1209713.5, similar to gb:X07696 KERATIN, TYPE I CYTOSKELETAL
 15 (HUMAN);, mRNA sequence.

ACCESSION AA726611
 MID 92744318
 VERSION AA726611.1 GI:2744318
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 579)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)

TITLE
 JOURNAL
 COMMENT On May 5, 1995 this sequence version replaced gi:797825.

CONTACT: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:646057
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 440.

FEATURES

source
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /map="6"
 /clone="IMAGE:1209713"
 /clone_1lb="Stratagene mouse skin (#937313)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
 sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT 173 a 149 c 151 g 106 t
 ORIGIN

Query Match 49.8%; Score 21.4; DB 37; Length 579;

Best Local Similarity 46.5%; Pred. No. 1.7e+02;
Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTTCGNCARGARARATCCCAATCDATRTGRGA 43
DB 393 GGAGATGACTTATCTCAAGAGAACCAAGAGAGAAATGAAA 435

RESULT 13

AA726685 638 bp mRNA EST 02-JAN-1998
LOCUS AA726685
DEFINITION v093h06.t1 Stratiogene mouse skin (#937313) Mus musculus cDNA clone

*IMAGE:1209755.5 similar to gb:X07686 KERATIN, TYPE I CTOSKELETAL
15 (HUMAN); gb:M1805 Mouse type I epidermal keratin mRNA, clone
pSNC-50, 3' (MOUSE);, mRNA sequence.

ACCESSION AA726685.1 GI:2744392
NID 92744392
VERSION AA726685.1
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 638)

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

THE MASHU-HHMI MOUSE EST PROJECT
Unpublished (1996)
On May 5, 1995 this sequence version replaced gi:798153.

TITLE JOURNAL
COMMENT

Contact: Marra M/Mouse EST Project
Mashu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:666099
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 445.

FEATURES
Source
Location/Qualifiers
1..638
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1209755"
/clone_1lb="Stratiogene mouse skin (#937313)"
/sex="females"
/tissue-type="whole skin"
/dev-stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1;
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3''

BASE COUNT 175 a 180 c 176 g 107 t

Query Match 49.8%; Score 21.4; DB 37; Length 638;
Best Local Similarity 46.5%; Pred. No. 1.8e+02;
Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTTCGNCARGARARATCCCAATCDATRTGRGA 43
DB 218 GGAGATGACTTATCTCAAGAGAACCAAGAGAGAAATGAAA 260

RESULT 14
LOCUS AA795990/c
DEFINITION AA795990 600 bp mRNA EST 09-FEB-1998
v066d10.t1 Stratiogene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:1151251.5 similar to gb:X07686 KERATIN, TYPE I CTOSKELETAL
15 (HUMAN);, mRNA sequence.

ACCESSION AA795990.1 GI:2858945
NID 92858945
VERSION AA795990.1
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 600)

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

THE MASHU-HHMI MOUSE EST PROJECT
Unpublished (1996)
On Jan 17, 1998 this sequence version replaced gi:2044105.

TITLE JOURNAL
COMMENT

Contact: Marra M/Mouse EST Project
Mashu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:624459
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 364.

FEATURES
Source
Location/Qualifiers
1..600
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1151251"
/clone_1lb="Stratiogene mouse skin (#937313)"
/sex="females"
/tissue-type="whole skin"
/dev-stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1;
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3''

BASE COUNT 119 a 143 c 170 g 168 t

Query Match 49.8%; Score 21.4; DB 38; Length 600;
Best Local Similarity 46.5%; Pred. No. 1.7e+02;
Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTTCGNCARGARARATCCCAATCDATRTGRGA 43
DB 112 GGAGATGACTTATCTCAAGAGAACCAAGAGAGAAATGAAA 70

RESULT 15
LOCUS C88410
DEFINITION C88410 450 bp mRNA EST 05-MAR-1998
C88410 Carp alignment-induced 48 hr cDNA Cyp17a1a cDNA clone
M131, mRNA sequence.

ACCESSION C88410
 MID 92924692
 VERSION C88410.1 GI:2924692
 KEYWORDS EST.
 SOURCE common carp.
 ORGANISM Cyprinus carpio

REFERENCE 1 (bases 1 to 450)
 Fujiki, R., Nakao, M., Shin, D. and Yano, T.
 Common carp genes differentially expressed in sodium alginate-elicited peritoneal exudate cells are identified by suppression subtractive hybridization technique unpublished (1998)

JOURNAL COMMENT On Apr 14, 1993 this sequence version replaced gi:639236.

CONTACT: Kazuhiro Fujiki
 Laboratory of Marine Biochemistry, Faculty of Agriculture
 Kyushu University
 Hakozaki 6-10-1, Higashi-ku, Fukuoka City, Fukuoka 812-8581, Japan
 Tel: 81-92-642-2896
 Fax: 81-92-642-2894
 Email: fujiki@agr.kyushu-u.ac.jp.
 Location/Qualifiers

FEATURES
 source
 1..450
 /organism="Cyprinus carpio"
 /db_xref="taxon:7962"
 /clone_lib="Carp alginate-induced 48 hr cdna"
 /cell_type="peritoneal exudate cells"
 BASE COUNT 104 a 117 c 119 g 110 t
 ORIGIN

Query Match 49.8%; Score 21.4; DB 39; Length 450;
 Best Local Similarity 48.8%; Pred. No. 1.7e+02;
 Matches 20; Conservative 11; Mismatches 10; Indels 0; Gaps 0;
 QY 2-GNATGAYTTCGNCAGARARARTGCCRAATCDATRTGRA 42
 Db 140 GACAGATTTCGCTCCACGACATGCCAAGTGAGTGTGA 180

Search completed: September 13, 1999, 16:14:04
 Job time: 5463 sec